University of Helsinki
Faculty of Educational Sciences
Helsinki Studies in Education, number 114

Satu Koivuhovi

STUDYING IN A CLASS WITH A SPECIAL EMPHASIS

Changes in children's competence beliefs and mathematical thinking skills and the role of social comparisons

Academic dissertation to be presented, with the permission of the Faculty of Educational Sciences at the University of Helsinki, for public examination in the Language Centre, Festive Hall, Fabianinkatu 26, on 17 Thursday June, 2021 at 4 pm.

Helsinki 2021

Pre-examiners
Professor Lars-Erik Malmberg, University of Oxford
Professor Piia Seppänen, University of Turku
Professor Todd D. Little, Texas Tech University (TTU)

Custos
Professor Risto Hotulainen, University of Helsinki

Supervisors
Professor Mari-Pauliina Vainikainen, Tampere University
Dr.Soc.Sci. Mira Kalalahti, University of Turku.

Opponent
Professor Todd D. Little, Texas Tech University (TTU)

Cover image
Satu Koivuhovi

The Faculty of Educational Sciences uses the Urkund system (plagiarism recognition) to examine all doctoral dissertation.

Unigrafia, Helsinki 2021

ISBN 978-951-51-7353-9 (paperback)
ISBN 978-951-51-7354-6 (PDF)
ISSN 1798-8322
ISSN 2489-2297

University of Helsinki, Faculty of Educational Sciences
Helsinki Studies in Education, number 114

Satu Koivuhovi

Studying in a class with a special emphasis
Changes in children's competence beliefs and mathematical thinking skills and the role of social comparisons.

Abstract
The purpose of this dissertation was to examine whether studying in a selective class with a special emphasis influences children's competence beliefs or mathematical thinking skills. The original idea behind this thesis was to explore empirically the commonly held assumption that studying in a class with a special emphasis improves motivation.

This dissertation consists of four original empirical studies, all of which examined the research question from several perspectives. In this summary part of this thesis, these perspectives were summarized into four research questions. The first overall aim was to examine how children's competence beliefs and mathematical thinking skills develop during the comprehensive school years. The second overall aim was to examine, how pupils who study in classes with a special emphasis differ from pupils who study in classes without a special emphasis in terms of background factors. The third overall aim, focused on the differences between classes with and without a special emphasis in the development of competence beliefs and mathematical thinking skills. Finally, the last overall aim explored more specific peer effects (i.e., the Big-Fish-Little-Pond effect, the Reflected Glory effect and the Peer Spillover effect) related to selective classes and explored whether these effects were visible in the Finnish context.

The data were drawn from a longitudinal learning-to-learn study in which the cognitive capabilities and motivational beliefs of 744 first graders were assessed and followed up throughout the comprehensive school years. Data consisted of several measurement points and tasks. The original sample size was increased during the follow up as children changed schools and new schools and classes were added to the sample. For the purpose of this thesis, data drawn predominantly from grades four to six (n=1025) and from seven to nine (n=2339) were used. Data were analyzed with statistical methods including single and multilevel structural equation models as well as repeated measures and analyses of variance.

The first finding of this thesis considered the overall development of children's competence beliefs and mathematical thinking skills and supported findings of prior research showing a decline in positive self-beliefs during the school years.

Decline in the competence beliefs was detected at both primary school and lower secondary school but the trajectories of different belief types varied. Children's mathematical thinking skills progressed, as expected, during the followed-up years.

Additionally, the second finding of this thesis confirmed findings of prior research and showed that classes with a special emphasis clearly differed from classes without a special emphasis in terms of pupils' background. Children who studied in classes with a special emphasis came from more highly educated families and had a higher grade point average (GPA) than pupils who studied in classes without a special emphasis. There were also detailed differences between classes with a special emphasis in terms of pupils' background.

The main results of this study considered the development of competence beliefs and mathematical thinking skills between classes with and without a special emphasis and showed interestingly, that there were no differences in the development. Even though pupils differed from each other initially due to the selective process of classes with a special emphasis, most of the differences in the development of competence beliefs and mathematical thinking skills were explained by these initial differences and the development was similar in different types of class after the background variables had been taken into account. Therefore, the results of this thesis gave no evidence of the assumed beneficial effects of emphasized teaching.

On the contrary, findings regarding the peer effects explored showed that the Big-Fish-Little-Pond effect was visible in the Finnish context whereas other peer effects were not. In other words, the results showed that the average achievement level of class predicted individual pupil's academic self-concept negatively. Therefore, these findings suggested that instead of the assumed beneficial motivational effects, studying in a highly selective class may have detrimental effects on individual pupil's self-beliefs.

Keywords: Classes with a special emphasis, selective classes, competence beliefs, action-control-beliefs, academic self-concept, mathematical thinking skills, longitudinal study, class composition, Big-Fish-Little-Pond

Helsingin yliopisto, Kasvatustieteellinen tiedekunta
Kasvatustieteellisiä tutkimuksia, numero 114

Satu Koivuhovi

Oppilaiden kompetenssiuskomusten ja matemaattisten ajattelutaitojen kehitys painotetun ja yleisopetuksen luokilla

Tiivistelmä

Tämän väitöstutkimuksen tavoitteena oli selvittää, miten painotetun opetuksen luokalla opiskelu näkyy lapsen oppimisessa, erityisesti itseä ja oppimista koskevien kompetenssiuskomusten sekä matemaattisten ajattelutaitojen kehityksessä. Tutkimuksen lähtökohtana oli tutkia empiirisesti yleistä painotettuun opetukseen liittyvää uskomusta, jonka mukaan painotetun opetuksen luokalla opiskelusta olisi hyötyä lasten oppimismotivaatiolle.

Tutkimus pohjautuu neljään empiiriseen osatutkimukseen, jotka kaikki tarkastelivat edellä mainittua teemaa omista näkökulmista käsin. Tässä tutkimuksen yhteenveto-osiossa nämä näkökulmat on tiivistetty neljäksi tutkimustavoitteeksi, joihin tässä yhteenveto-osiossa vastataan. Yhteenveto-osion ensimmäinen tutkimustavoite liittyi lasten kompetenssiuskomusten ja matemaattisten minäkäsitysten kehityksen tarkasteluun peruskoulun aikana. Toinen tavoite oli selvittää, minkälaisia lähtökohtaisia eroja painotetun ja yleisopetuksen luokilla opiskelevien oppilaiden välillä oli taustatekijöiden suhteen tarkasteltuna. Tämän jälkeen keskityttiin tarkastelemaan, miten oppilaiden kompetenssiuskomukset ja matemaattiset ajattelutaidot kehittyivät painotetun ja yleisopetuksen luokilla. Viimeiseksi tarkasteltiin vielä erikseen aiemmissa tutkimuksissa usein selektiivisiin luokkiin liitettyjä vertaisryhmävaikutuksia (Big-Fish-Little-Pond-Efekti, Reflected Glory-efekti ja Peer Spillover-Efekti) ja sitä, oliko nämä efektit havaittavissa suomalaisessa kontekstissa.

Tutkimuksen aineistona toimi oppimaanoppimisen pitkittäistutkimus, jossa 744 ekaluokkalaisen kognitiivisia taitoja sekä motivationaalisia uskomuksia arvioitiin peruskoulun ensimmäiseltä luokalta viimeiselle. Aineistoa kerättiin useissa eri mittausajankohdissa ja monilla eri mittareilla. Alkuperäisen aineiston otoskoko kasvoi tutkimuksen edetessä kun seurantaan otettiin lisää oppilaita alkuperäiseen otokseen kuuluneiden oppilaiden vaihtaessa luokkaa tai koulua. Tässä tutkimuksessa hyödynnettiin erityisesti aineistoa vuosiluokilta neljästä kuuteen (n=1025) sekä seitsemästä yhdeksään (n=2339). Aineisto analysoitiin tilastollisilla menetelmillä, jotka pitivät sisällään erilaisia yksitasoisia ja monitasoisia rakenneyhtälömalleja.

Tutkimuksen tuloksena todettiin, että kuten aikaisemman tutkimuksen perusteella osatiin odottaa, kaiken kaikkiaan lasten kompetenssiuskomukset

heikkenivät kouluvuosien aikana. Lasku positiivisissa kompetenssiuskomuksissa havaittiin sekä alakoulua että yläkoulua tarkastelevissa analyyseissa, mutta eri uskomustyyppien kehityskaarissa oli eroja. Lasten matemaattiset ajattelutaidot kehittyivät odotetusti seurannan aikana.

Myös toinen tämän väitöstutkimuksen päätuloksista vahvisti aikaisempien tutkimusten tuloksia ja osoitti, että painotetun opetuksen luokilla opiskelevat lapset erosivat yleisopetuksen luokilla opiskelevista lapsista taustatekijöiden mukaan tarkasteltuna. Painotetun opetuksen luokilla opiskelevat lapset tulivat korkeammin koulutetuista perheistä ja menestyivät koulussa lähtökohtaisesti paremmin kuin yleisopetuksen luokilla opiskelevat lapset. Myös eri painotusalojen välillä oli eroja oppilaiden taustatekijöissä.

Tämän väitöstutkimuksen päätavoite oli tarkastella kompetenssiuskomusten ja matemaattisten ajattelutaitojen kehitystä painotetun opetuksen ja yleisopetuksen luokilla. Keskeisimpänä tuloksena voidaan todeta, että suurimmaksi osaksi eroja kehityksessä ei ollut. Vaikka painotetun opetuksen ja yleisopetuksen oppilaat erosivat toisistaan lähtökohtaisesti taustatekijöiden suhteen, oli kompetenssiuskomusten ja matemaattisten ajattelutaitojen kehitys suurimmaksi osaksi samankaltaista erilaisilla luokilla ja taustatekijöissä olleet lähtökohtaiset erot selittivät valtaosan havaituista eroista luokkien välillä. Näin ollen tutkimuksen tulokset eivät tukeneet ennakko-olettamuksia painotetun opetuksen positiivisista vaikutuksista oppilaiden motivaatioon. Päinvastoin, tutkimuksen tulokset antoivat tukea ns. Big-Fish-Little-Pond-efektille ja osoittivat, että tämä kansainvälisesti paljon tutkittu ilmiö oli näkyvissä myös suomalaisessa kontekstissa ja luokan keskimääräinen osaamistaso ennusti negatiivisesti oppilaan minäkäsitystä. Näin ollen tutkimuksen tulokset osoittivat, että oletettujen positiivisten vaikutusten sijaan valikoidussa vertaisryhmässä opiskelu voi itse asiassa heijastua negatiivisesti yksilön minäuskomuksiin.

Avainsanat: Painotetun opetuksen luokat, valikoivat luokat, kompetenssiuskomukset, akateeminen minäkäsitys, matemaattiset ajattelutaidot, pitkittäistutkimus, luokkakompositio, Big-Fish-Little-Pond-efekti

Dedicated to my grandparents Kaija and Santtu, who always asked me what I did there at the university and who would have been so proud of me now.

ACKNOWLEDGEMENTS

Finally, I have reached this fortunate position when this never-ending project has come to the stage that only the final words of this book need to be written. I feel that this is also the best part as over the years there have been so many wonderful people around me and so I am really happy that now I have the opportunity to thank you all.

I got my first glimpse of the stunning world of research in the course on quantitative methods in my early years at the University of Helsinki. At the start of that course, I had no idea what quantitative methods meant. I remember that I thought that the course name sounded a little bit like something that maybe Harry Potter could study in his wizard school, which probably explains how little I knew about statistics or quantitative methods before that. Also, my understanding about research in general was very limited. However, during that course, I felt that a whole new world opened up for me. The course was taught by two wonderful "quantiwizards" Leila Pehkonen and Ilse Erikson, who had expertise about methods, but also magnificent pedagogical skills, which made the learning experience inspiring.

The idea of studying for a doctorate started to evolve in my mind while I was writing my Bachelor's thesis. I remember how mind-blowing it was to run an exploratory factor analysis in SPSS for the first time, and get some results from the self-collected data. Professor Markku Niemivirta who taught the bachelor's seminar was not only an expert in the field of quantitative methods, but also an exceptional teacher who shared my feelings of excitement, and he patiently answered my multiple emails between the seminar meetings. During that seminar, I quickly understood how little I knew, but also that I would definitely want to learn more.

After the Bachelor's degree, I went on to the Master's programme for educational sociology and policy (KSP) led by Professor Hannu Simola. I was interested in sociology, but in addition, the programme was quite research-oriented, and I thought that it could help me figure out whether the idea of undertaking a PhD would make any sense. Afterwards, I can say that it did, in many ways. Inspiring discussions with Hannu and his fatherly guidance and genuine caring for us all during the years at KSP was something that I will never forget. I received extensive advice during those times, which will guide me in the future after this dissertation, and Hannu's winged words such as "Only good players can be lucky" have always echoed in my mind, when I have written grant proposals.

During the KSP years, I got a chance to see how a research project works, while I was working as a research assistant for VAKOVA and later PASC. During those years, the idea of postgraduate study grew stronger, and I started to think of that as an option. While I was finishing my Master's thesis, I spoke with our amanuensis Tuomo Aalto about my future options, and he advised me to speak with Professor Jarkko Hautamäki, who was the founder and at that time the leader of the Centre for Educational Assessment (CEA) at the University of Helsinki. After that discussion with Tuomo, I went directly to knock on Jarkko's door (thinking that I could book an appointment for him later in the future). Surprisingly, Jarkko asked me to sit down and say what I had to say. I remember that I was so nervous and mumbled something about my plans, but after the meeting, Jarkko welcomed me and said, "We will figure out something for you!" And that was the beginning of my PhD project.

So, now, finally, to the actual thank-you part…

First, I want to express my gratitude to the pre-examiners of this thesis, Professor Lars-Erik Malmberg, Professor Piia Seppänen, and Professor Todd D. Little, for all the insightful and valuable comments, which improved my thesis in its final stage. I feel deeply honoured that Professor Little promised to serve as an opponent in the public defence of this thesis. Even though I feel extremely nervous and excited about the upcoming event, at the same time, I am eagerly waiting for it, as it will be such a privilege to have the opportunity to discuss my work. Professor Little, I appreciate your expertise in the field of educational psychology and quantitative methods, and having a chance to discuss with you (even online) will be a big learning opportunity for me. So, thank you in advance for it. I hope that in the future, I might have a chance to meet you in person.

In addition, I am most indebted to the co-authors of my fourth article, Professor Herb Marsh, Dr Theresa Dicke, Associate Professor Baljinder Sahdra, Dr Jiesi Guo, and Professor Philip Parker. If someone would have told me when I started my PhD and read some first papers on BFLPE that during these years I would have a chance to collaborate with you, I would never had believed it! It has been such a huge honour to meet you and work with you! Especially now in these days of lockdown, I remember my visit to Sydney as a miraculous journey during which I learned so much. Seeing the way you work, and having a chance to collaborate with you has been so inspiring and exciting that I honestly don't have words to describe what it means for me! I hope that in the future, there will come a day when it will be possible to travel and meet you all again. Special thanks must be aimed at Herb. It feels as though I had the opportunity to rock with the biggest rock star ever, and it is sometimes a little difficult for me to understand that it actually happened! Your expertise is outstanding, and the way you encouraged me kindly and patiently was something extraordinary. In addition, Theresa, thank you

so much for all the advice and support during our collaboration! You always had the patience to explain things to me over and over again, and there were no questions I would not have been able to ask. I am looking forward to the day we can meet again!

Professors Emeritus Hannu Simola and Jarkko Hautamäki. Two Grand Old Men of Science. Throughout the years, I have enjoyed the discussions with you and tried to absorb wisdom from you. I feel that I have learned so much from you both regarding scholarship and life in general. It has been so inspiring to discuss with you and listen to your lectures and speeches on several occasions. They have always given me new things to consider and challenged my thinking.

At the beginning of this project, I also received plenty of help and inspiration from Professor Markku Niemivirta. I will be forever grateful for all the help you gave me while I was writing the first research plan of this thesis. I admire your expertise on the field of motivation research as well as on quantitative methods, and it was a privilege to have a chance to learn and collaborate with you while writing the first article of this thesis. Thank you!

My biggest thanks go to my wonderful supervisors Professor Mari-Pauliina Vainikainen and Docent Mira Kalalahti. This project has been a long roller coaster ride with many ups and downs. I know that I was probably not the easiest person to supervise, as I don't always listen to all instructions, and sometimes I can be quite stubborn. Especially at moments of disappointment (such as rejections or bad reviews), I tend to be quite emotional, and I don't know how many times I declared to you that I was done with this thesis. But repeatedly, you were supportive and encouraging, and you believed in me and I am so grateful to both of you for these years.

Martsu. After I had knocked on Jarkko's door, you were the first person he introduced me to at CEA. You welcomed me to CEA so warmly that I immediately felt at home. You were still finalising your own PhD when we met, and since then, it has been a pleasure to see how you started your own academic career and became a professor. Doing a PhD under your supervision has been a joyful journey, and I have many good memories of the moments when we sat in your room in Siltavuorenpenger studying new analyses or discussing the interpretations of some results, with Jarkko drawing figures. Or moments when we picked mushrooms in the forests and spoke about the future or life in general. Even though you were busy, you always had time for me and my questions. Whether it was about a theoretical, methodological, or a personal problem, you were always able to help me. Thank you for all these years! I am so happy that I have had the privilege to get to know you and become your friend. You are simply awesome!

Mira. You were my supervisor during my Master's thesis, so we have quite a long history together. So, you knew probably pretty well what to expect, when you committed to guiding me during this project. I feel that I don't have enough

words to thank you for all the support and guidance I received from you during these years. You have always been the calm supporter who has had time to listen and discuss all the matters that have arisen. In addition, your comments and supervision of my writing over the years have always been exact and sharp, and made me elaborate my ideas. You and Janne were also Tommi's supervisors, so I bet it will be a relief for you to get rid of us ☺ However, I hope that we will still have many projects ahead of us in which our paths will cross. Thank you both so much!

In addition, I want to thank all the people I met and worked with at the university. Thank you, all the wonderful people in KUPOLI, with whom I got a chance to work during the years of KSP/VAKOVA/PASC and later. Thank you Eeva Rontu, Hannele Pitkänen, Antti Paakkari, Heidi Vartiainen-Huilla, Saija Volmari, Sonja Kosunen, Jaakko Kauko etc. for all the discussions and evening get-togethers over the years.

The most important academic home for me during these years has been the Centre for Educational Assessment (CEA) at the University of Helsinki, and later, also Martsu's research group (REAL) in Tampere. Thank you, all the amazing people for your existence! When I think of my years in CEA, I remember most clearly all the moments of laughter and joy, lunch and coffee breaks full of jokes, or situation-based comedy. For that, for producing the wonderful atmosphere of CEA, I want to thank all the CEA people, especially Jarkko Hautamäki, Mari-Pauliina Vainikainen, Raisa Ahtiainen, Ninja Hienonen, Meri Lintuvuori, Sirkku Kupiainen, Risto Hotulainen, Helena Thuneberg, Natalija Gustavson, Marja-Liisa Kieksi, Irene Rämä, Sanna Oinas, Laura Kortesoja, Lauri Heikonen, Laura Nyman, Esko Lindgren, Jukka Määttänen, Mikko Asikainen & Jukka Marjanen (I will always count you as one of the CEA people, as you were there when I started). Thank you for all the lovely moments of laughter. I hope that in the future, post-pandemic world, we will be able to meet more in person, I miss that so much!

Some special thanks to:

Risto. Thank you for acting as the custos for my doctoral defence. In a time when everything, even doctoral defences, have to take place online and at distance, it will be nice to be able to have at least you there, so that everything is not happening remotely. I am sure that your calm presence in the defence situation will help me to keep my nerves!

Sirkku. Thank you for all the insightful comments and conversations during the years, especially regarding learning-to-learn, school choices and differences between classes. At the beginning of this project, I remember our Friday morning discussions with you, Martsu, and Make, and how many ideas and inspiration I

received from them. I truly value your expertise and straight way to give feedback, and discussions with you have always given me new perspectives to consider.

Helena. I want to thank you for all the emotional and professional support, which you have given me in times when I needed it. I remember crying with exhaustion when I tried to understand multilevel models after Mai had just been born and was with me at the University, and your support, help, and kind words during those moments were invaluable. Later, I also enjoyed discussions with you about more theoretical things, and those discussions are something that I hope we can continue in the future.

Natalija. Thank you for all the support and for the motherly care during these years. You have always encouraged me and cheered me up to continue, even when it felt difficult. In addition, I want to thank you for being the best research secretary in that longitudinal data collection from which my data comes. Without you, the data I have been using would entail much more missing information. Your persistent effort in tracing the pupils who changed schools or classes over the years was exceptional.

Mikko. I must thank you for all those times when you helped me with technical problems or excel things or whatever. You are the best!! In the same vein, Jukka Marjanen, I could say the same words to you. I still have my orange notebook from my first years in CEA, which is full of the notes I made based on your advice of either Excel or SPSS (my Mplus and R notes from you are not in that same notebook as those I have nowadays in my computer!). So, thank you both for being my technical support (also Tommi, I must mention you here!! Thank you for all the tables and figures, which you taught me to do) but also for being such hilarious company in all the occasions during the years.

Laura Nyman. It has been so nice to get to know you and discuss with you about the theoretical issues regarding motivational constructs during the last year. I enjoyed this a lot, and I believe that we will have many shared projects ahead of us. We already have this wonderful new idea about the value of crying during the writing process, and I believe that we will cultivate it to become a new master theory in the field of motivation! ☺

Cristiana Mergianian, thank you for proof-reading my texts!

In addition, I have had many fellow students outside of the already mentioned groups with whom I have been able to share the best and worst parts of writing a PhD. It is impossible to name all of you, but if you read this and recognize yourself, feel yourself important and lauded!

The most important fellow sufferer of mine during these years has been Anna Rawlings. I am so happy that we have had the chance to share these years together, including these last months before our respective doctoral defences. I am so proud of us that we both did this and now we can finally have that glass (or glasses, or bottles) of sparkling wine we have long been waiting for. It has been a blessing to

have someone in the same situation with whom to share all the varied moments during this journey. We have both had our disappointments, and it has felt like you understand me even without words (certain emojis or GIFs can capture the emotions pretty well…). Now it is time to celebrate!!

In addition, I want to thank my friends and family outside academia.

Salla and Tiina. Thank you for your friendship. We have shared many unforgettable moments since we got to know each other in Liusketie at the edge of the sandbox. I am so happy that I have friends like you!

Salla. Your house has been always open to us, just like your heart. You are probably the most warm-hearted person I know, and you and your whole family are so important to me and my family. We have so many good memories of our trips together, and I can't wait to get into new ones. I hope it will happen soon.

Sonja. You have been my best friend since our lower secondary school years. Thank you for producing me the best drama outside academia and always being there for me. Even though we now live far away from each other, I have always felt that in a way you are close. I hope that in the future, we will have more opportunities to meet in person.

Anwar. This is the book I have been writing, and one of the reasons why I have been so busy. Thank you for your friendship, even though we rarely have time to meet. You have given me perspective to notice and appreciate my life in a way in which I think I would not have done without having the chance to meet and discuss it with you.

Jussi & Maria. Jussi, you have been in my life forever, and we have shared so many things together. I am thrilled and honoured to act as your best man in your upcoming wedding; I can't wait for that! I wish that you and Maria will live together happily ever after, I am so happy for you <3

Ali Shah. Thank you for being a friend for me and my whole family. *Kiitos, kun olet tullut meidän perheen ystäväksi. Kiitos kaikista kerroista, kun hoidit Joelia, että minä pystyin tehdä töitä. Toivottavasti tulevaisuudessa ehditään nähdä enemmän.*

To my mum and dad, Mari & Risto Koivuhovi, thank you for being my parents and giving me good opportunities for life. I have always felt myself loved by you, and you have supported my choices in my life regardless of how rational or irrational they have been. Thank you, Niko, for being my best (and only) brother! You have offered some great peer support for processing childhood traumas together ☺ In addition, Ralph and Eija, thank you for being there for my mother and father after their divorce. I am happy that they have both met a nice person with whom to share life together. Thank you all (including Mummi, Nonno, and None) for being part of my children's life too!

Tommi's parents Anu and Jussi. Thank you for helping us with kids and the never-ending renovation project called "our house". Your help has been priceless!

Kirsi & Kari, thank you for being my godparents and present in my life throughout the years. Thank you for taking care of our children and of course also Carla!

Aila & Arto, thank you for welcoming us to Pielavesi every summer, moments there have been unforgettable and something to look forward to during winters.

Leena, thank you for being part of our life and taking care of Mai especially. I hope that in the future, we will again have more time to spend together.

Sharif. Thank you for becoming part of our family. I am so happy that you got the chance to stay in Finland. It makes me so happy to see how you have been able to settle down and start to build your life here, even though it most certainly has not been easy for you.

Families of Janne and Johannes, thank you for being part of our life. Thank you Vallilan mummo Pirkko and Jussi and Katja & Tommi for being there for my children and for us during these years on a range of occasions.

One of the most difficult parts in this project, maybe even more difficult than all the complicated statistical analyses, has been balancing between my own research and my wonderful family. Even though it has been quite hard and sometimes even physically exhausting, at the same time, these years have been unforgettable and full of life and joy. I am so grateful for my four marvellous children, Julius, Mio, Joel and Mai. You are my everything. I hope that you will have the courage to follow your dreams in life and will have many amazing adventures ahead of you. I love you. I know that you are pretty used to me studying, as it has been like that for all the years that you have existed. I started studying when you, Julius, were born, did my Bachelor's thesis just before you, Mio, arrived, and Joel, you were born when I had just started my doctorate in the position at SEDUCE. And last but not least, you, Mai, rushed into this world (more precisely into the backseat of Grandmas aka "*mummin auton värinen*" car) after our Australian trip. So, you are all used to me always being a little unfocused or absorbed with something related to my studying, sometimes more and sometimes less. However, I promise you that for this summer, after my dissertation, I will try to clear my head as best as I can from all the work and research issues, and focus solely on you and the summer. So, boys, I promise you that it will be as much *Pesäpallo* as possible (every possible *Superpesis ottelu* which we can go to see live together, in addition your own matches of course!! Sorry, Tommi!)

Tommi. We decided that we will do this together, and now we have done it! Time will tell what comes next, but I am thankful to you for all these years. I feel sad that during these years, we have had so little time for the two of us, and we have kind of lost ourselves into this chaos of life. Despite all the obstacles and difficult times, we decided to stick together, and I am happy for all the great

memories and adventures we have had together, both with the children and sometimes (years ago) also just you and I. I hope that in the future, with both our theses behind us, we could have more time for each other. Even though our life sometimes feels like a roller-coaster ride, I still love you, at least every now and then. Thank you also for being my personal assistant; some of the tables in my articles would not have been so elegant without you, even though I must say that I have learned a lot (see, for example, Figure 3!).

Finally, I want to thank all the anonymous people who had a role in this process:

Participants in my research: children, parents and teachers. Thank you for taking part in the data collection. Without you, this research would not have been possible!

Peer reviewers of the original articles: Thank you for giving me a lesson in how to cope with disappointments. I understand that you have an important and often unappreciated role, and I must admit that sometimes I hated you. However, now I say thank you! Your valuable comments helped me to improve the original articles, and they would not be what they are without you.

Thank you also SEDUCE doctoral programme for providing me with funding during these years and enabling me to focus on research full time.

In Helsinki, remotely from our home (alone!), 20 May, 2021

Satu Koivuhovi

Contents

List of tables

List of figures

List of Appendices

ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their roman numerals (Studies I-IV).

Study I Koivuhovi, S., Vainikainen, M.-P., Kalalahti, M. & Niemivirta, M. (2019). Changes in Children's Agency Beliefs and Control Expectancy in Classes With and Without a Special Emphasis in Finland from Grade Four to Grade Six. *Scandinavian Journal of Educational Research,* 63 (1), 427–442. https://doi.org/10.1080/00313831.2017.1402364

Study II Koivuhovi, S., Vainikainen, M.-P. & Kalalahti, M. *(2020*). The effect of Studying in Selective Classes on the Change of Pupils' Action-Control Beliefs during Lower Secondary School in Finland. *Scandinavian Journal of Educational Research.* Advanced online publication. https://doi.org/10.1080/00313831.2020.1833246

Study III Koivuhovi, S., Vainikainen, M.-P. & Kalalahti, M. (*2021*). Oppilaiden matemaattisten ajattelutaitojen ja matematiikkaminäkäsityksen kehitys painotetun opetuksen ja yleisopetuksen luokilla neljänneltä luokalta kuudennelle. [The development of pupils' mathematical thinking skills and mathematical self-concept in classes with and without a special emphasis from fourth to sixth grade] *Kasvatus,* 52 (1), 22–36.

Study IV Koivuhovi, S., Marsh, H.W., Dicke, T., Sahdra, B., Guo, J. Parker, P.D., & Vainikainen, M.-P. (*2020*). Academic Self-concept Formation and Peer-Group Contagion: Development of the Big-Fish-Little-Pond Effect in Primary-school Classrooms and Peer Groups. *Journal of Educational Psychology*. Advanced online publication. https://psycnet.apa.org/doi/10.1037/edu0000554

Author's contribution

This doctoral thesis has been conducted as a member of REAL, the Research Group for Education, Assessment and Learning, and consists of four articles, all of which have Satu Koivuhovi as the first author.

In the first article, Koivuhovi planned the research design with supervisors (i.e., Professor Markku Niemivirta, Professor Mari-Pauliina Vainikainen and Dr Mira Kalalahti), performed the analyses and had the main responsibility of writing and submitting the article. Supervisors commented on the analyses and text throughout the writing process. Professor Markku Niemivirta was especially responsible for helping Koivuhovi with the analyses and he also revised the first draft of the manuscript. Later revisions were made with Koivuhovi and all the supervisors.

In the second and third articles Koivuhovi planned the research design, performed the data analyses and took the main responsibility of writing, submitting and revising the articles. Professor Mari-Pauliina Vainikainen was especially responsible for helping Koivuhovi with the analyses and Dr Mira Kalalahti commented and revised manuscripts throughout the writing processes.

In the final article, Koivuhovi served as a first author and Professor Herbert Marsh was the corresponding author. The research design was planned together and Koivuhovi completed all the preliminary analyses and the first draft of the manuscript. Professor Marsh commented and gave advice throughout the process and revised the final version of the manuscript. In addition, other authors of the manuscript commented on the analyses and text throughout the writing process. Dr Theresa Dicke was especially responsible for helping Koivuhovi with the data analyses. Associate Professor Baljinder Sahdra performed all the preliminary data analyses with sociometric data so that it was in the right format for the actual analyses.

1 INTRODUCTION

Tracking pupils based on their abilities or other aptitudes is common in many countries. Various allocation and grouping practices can be seen as an attempt to manage the diversity and individual differences in education and a way to optimize teaching to give maximum benefits for all pupils (Dupriez et al., 2008; Gottfried, 2014). The effects of tracking on individual pupils' achievements or motivation have been studied extensively and the question of whether or not to track pupils is an on ongoing debate in educational research (e.g., Hattie, 2002, 2009; Loveless, 2009; Zimmer, 2003; Dracup, 2014; Steenbergen-Hu et al., 2016). Arguments favoring tracking claim that allocating pupils into homogenous learning groups makes it possible to adjust teaching to meet each pupil's individual needs, and therefore it is believed that tracking would be of benefit to all (Tieso, 2003). However, opponents emphasize the negative equality effects of tracking and argue that benefits are not equally distributed (Hanushek & Wößmann, 2006; Hattie, 2002).

The basic education system in Finland has often been presented as an example of a non-tracking system (Wößmann, 2009). Therefore, the top results that Finland has received in many international evaluations (e.g., PISA, TIMMS) can be seen to speak against tracking. The example of Finland demonstrates that top learning results can be achieved in a comprehensive school system in which all children study according to the same national core curriculum. However, the Finnish basic education system has one unique feature that stands out from the otherwise comprehensive system. Since the changes in education legislation in the 1990s, schools have had the chance to specialize in certain subjects (e.g., in music, languages etc.) by offering emphasized teaching (Seppänen, 2003; Seppänen, 2006; Kalalahti et al., 2016). Schools could use aptitude tests to select pupils in emphasized teaching and teaching can be organized in their own classes i.e. *classes with a special emphasis* (Seppänen, 2003; Simola et al., 2015).[1] Since their establishment, classes with a special emphasis have become popular, especially in urban Finland, in fact, over 30% of all pupils in some cities study in this type of class (Simola et al., 2015; Seppänen, Kalalahti, et al., 2015). Therefore, classes with a special emphasis have been called "implicit tracks" within the Finnish comprehensive school and they have been a hot topic in educational political discussion (Berisha & Seppänen, 2017; Varjo & Kalalahti, 2019). Classes with a special emphasis have been shown to attract especially well-achieving children

[1] Educational legislation speaks only about emphasized teaching and schools have right to organize the emphasized teaching either in separate classes or withing general classes (Seppänen, 2003). Even though there are no exact statistics of how different municipalities organize emphasized teaching in practice, it seems that the most common way to organize it is in separate classes (Simola et al., 2015, see also Table 1 in Chapter 3.2.1).

from affluent families and therefore, it has been argued that they endanger the equality of the Finnish comprehensive school as they differentiate children's school paths according to their family background (Kalalahti et al., 2015; Varjo & Kalalahti, 2019). However, politicians endorsing emphasized teaching and school choice have seen that selectivity based on pupils' own interest could be important in ensuring the individuality of teaching (Seppänen, 2003; Silvennoinen et al., 2015). It is also believed that having the right to choose a class with a special emphasis according to an individual pupil's interest, could increase motivation to do schoolwork. In particular, parents seem to believe that the atmosphere in these selective classes would enhance children's motivation and enthusiasm for schoolwork and thus be of benefit to the child (Kosunen & Carrasco, 2016; Kosunen & Seppänen, 2015).

Despite many assumptions and speculations regarding the effects of emphasized teaching, empirical studies analyzing such effects have been scarce. Prior studies have focused mainly on describing the selection process in families (e.g., Kalalahti et al., 2016; Kosunen, 2014), but the actual effects that studying in a class with a special emphasis may have on an individual pupil's motivation or learning have remained unstudied. Therefore, the aim of this dissertation is to examine how children's competence related beliefs and mathematical thinking skills develop in classes with and without a special emphasis. Different aspects of this phenomenon were studied in four original empirical articles[2], which are based on a longitudinal study spanning the comprehensive school years. In this summary, this main research problem will be addressed in two phases: first by summarizing the results of prior research in tracking and peer effect literature in Chapter 2.4 and second by summarizing the results of the original articles in Chapter 4. Before proceeding to the tracking literature, the theoretical framework of competence beliefs will be presented in Chapters 2.1-2.2. Next, the focus will shift to different tracking practices (Chapter 2.3), followed by the Finnish "tracking model" in classes with a special emphasis in Chapter 2.5. Chapter 2.6 completes the theory part of this summary and ties together the concepts and the research design of this study. After that, detailed information about the aims and methods of the present study, including information about data and measures, are presented in Chapter 3. After the summary of the original results in Chapter 4, the implications of the results will be discussed in Chapter 5.

[2] This dissertation is based on four empirical articles, which have been published in national and international peer-reviewed journals. All the original articles explored the main research question of this thesis from different perspectives and used longitudinal learning-to-learn data covering all basic education years (see Chapter 3.2). In study I, the focus was on children's action-control beliefs at the primary school level. Study II continued within the same theoretical framework but explored those beliefs at the lower secondary school level. Study III focused on children's academic self-concept and mathematical thinking skills at the primary school level and study IV continued by exploring more specific peer effects on academic self-concept at the primary school level. Copies of these original articles have been included after the summary part of this thesis.

2 THEORY

2.1 Competence beliefs as measures of motivation

Motivation is a concept which is often used in everyday language to describe the reasons or intentions for actions. Even though there seems to be a shared understanding about the meaning of the concept in everyday language, in research it has been operationalized and used in multiple ways. The shift from behavioristic simplistic stimulus-reaction models to cognitive models, in which behavior is understood more broadly, also changed the focus in motivation research (Pintrich & Schunk, 1996; see also White, 1959). Cognitive approaches stressed the importance of cognitive processes in mediating the influence of external factors on human functioning (Bandura, 1999). Instead of focusing on analyzing the question of what makes a human act, the focus shifted to the directionality of action (Pintrich & Schunk, 1996). In modern motivation theories, the focus has therefore been on beliefs, values and goals of action and relations between them (Eccles et al.,1998; Eccles & Wigfield, 2002).

Many theories of motivation make a distinction between beliefs that individuals have about themselves as an actor (probability or expectancy of success) and between the beliefs they have about the task they are faced with (importance and value) (e.g., Bandura, 1977; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). Motivation is thought to emerge from an interplay between both types of beliefs (Pintrich & Schunk, 1996).[3] In this dissertation, the focus has been on two types of self-beliefs: academic self-concept (e.g., Marsh et al., 1988; Marsh & Craven, 2006; Marsh & Martin, 2011; Shavelson et al., 1976) and action-control -beliefs (i.e. action-control theory by Skinner et al., 1988, 1990, 1998). Both fall into the first category in the above-mentioned classification and thereby deal with beliefs that individuals have about themselves as agents. These kinds of competence self-perceptions are central to many motivational theories as the feeling of competence has been understood to have an important influence on

[3] In their writings Eccles et al., (1998; see also Eccles & Wigfield, 2002) have divided motivational theories according to the key motivational questions, which they aim to answer. In addition to theories, falling into the category of competence beliefs and thus answering the question "Can I do this task?", other branch of theories aim to answer to the question, whether and how an individual is willing to engage into action. Therefore, this branch of theories includes self-determination theory (e.g., Deci & Ryan, 2000) and goal theory (for a review see Elliot & Hulleman, 2017). As a third category in their classification, Eccles et al., (1998; see also, Eccles & Wigfield, 2002) point out theories (e.g., Zimmerman, 1989), which connect motivation into cognitive functioning and therefore, examine for example those cognitive strategies and self-regulation skills, which translate the motivational beliefs into regulated behavior. Even though this categorization is a simplification and many theories overlap with the categories, I have found it to be a useful tool for navigating in the jungle of motivational theories.

human beings' social and emotional development (e.g., Bandura, 1999; Jacobs et al., 2002; Little et al., 2002; Muenks et al., 2018; Wigfield & Eccles, 2000). In the field of education, an individual's feeling of competence over his/her own performance has been seen to lead to many positive outcomes both in the present achievement situation but also in later educational career (Bandura, 2004; Bong & Skaalvik, 2003). Even though both academic self-concept and action-control beliefs can be classified under the term competence related self-beliefs and thus answer to the question "Can I do this task?" (Eccles et al., 1998; Eccles & Wigfield, 2002), their theoretical background is quite different. Academic self-concept focuses precisely on domain specific beliefs that individuals have about themselves in certain areas whereas action-control beliefs operate at a more general level and offer a comprehensive model of action.

2.1.1 Action-control beliefs

Action-control beliefs refer to beliefs formulated in a theoretical framework of action-control theory (Skinner et al., 1988, 1990, 1998). The theory stresses human beings' psychological need for control (Skinner, 1996; Skinner et al., 1998), which has been addressed in a broad range of theories (e.g., Bandura, 1977; Weiner, 1985; White, 1959; Deci & Ryan, 2000). The perceived control over occurrences is important, as it gives the individual a feeling of competence and therefore forms the basis for human beings' motivated action (Bandura, 1993, 1999; Skinner, 1996). Thus, in action-control theory, individuals' control expectancies are central.

In action-control theory, action is conceptualized as a threefold construct, encompassing the agent, aims and means of the action (see Figure 1). The individual, the agent, possesses different kinds of beliefs (i.e., *means-ends beliefs*, *agency beliefs* and *control expectancy*) concerning each part of the action and these beliefs shape the individual's actions and strivings. Each set of beliefs can be conceptualized and measured separately (Skinner et al., 1988, 1990), but together they comprehensively describe the individual's motivated action.

Means-ends beliefs (e.g., Little & Lopez, 1997; Roque et al., 2014; Skinner et al., 1988) refer to causality beliefs that individuals have about the means (i.e. effort, ability, luck) that are effective for achieving desired goals (such as learning or good grades). *Means-ends beliefs* differ from self-attributions (Weiner, 1985), as they are generalized perceptions that individuals have about the usefulness of certain means in relation to desired goals without a reference to the agent (Little et al., 1999; Little & Lopez, 1997). *Means-ends beliefs* can be divided into intra-agentic and extra-agentic means regarding whether they refer to means emerging from the agent (i.e. effort and ability) or external factors (i.e. luck) (Geldhof & Little, 2011; Little et al., 2002).

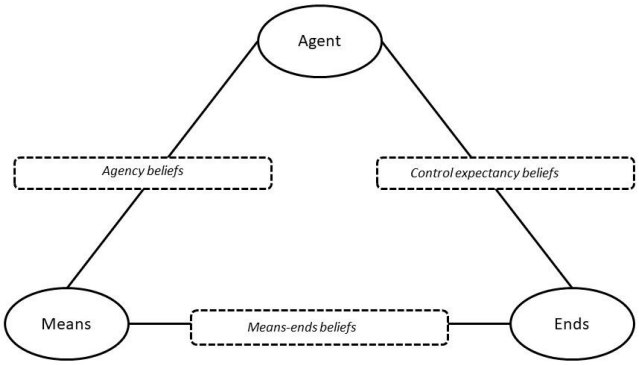

Figure 1 A schematic representation of three sets of beliefs in action-control theory about the relations among agents, means, and ends (modified from Skinner et al., 1988)

Agency beliefs are beliefs that the individual has about him/herself as an agent in relation to the possible means of action (Skinner et al., 1988, 1990). Thus, agency beliefs refer to the pupil's beliefs of him/herself as a learner (Little et al., 2001). Of all of the three belief types in the action control theory, *agency beliefs* have been noted as the best predictor of pupils' academic success (Little et al., 2001; Skinner et al., 1990).

The last set of beliefs, *control expectancy*, refer to the relation between the agent and the desired goals without a reference to specific means (Skinner et al., 1988, 1998). Therefore, they refer to the expectancy of the individual about his/her own opportunities to reach the desired goals. Thus, in the context of school, these beliefs comprise expectations of how likely one is to, for example, succeed at school or get good grades.

Motivated behavior stems from an interplay of all these beliefs, which have their own functions, but still work in relation to each other. Therefore, changes in one set of them is reflected in an individual's action (Malmberg & Little, 2007; Skinner et al., 1988). In other words, for example, if a pupil's means-ends beliefs of abilities are strong (i.e. s/he believes that abilities are an essential prerequisite of school success), the implications on achievement motivation depends on the agency beliefs of ability that this same pupil possesses (i.e. how able s/he sees her/himself). Studies have shown that action-control beliefs can predict academic goals (Lopez, 1999; see also Geldhof & Little, 2011) and mediate motivational styles (Walls & Little, 2005). In addition, studies have concluded that certain belief combinations seem to be connected with more positive development than others (Little et al., 2002; Walls & Little, 2005). For example, Malmberg and Little (2007) identified typical adaptive and maladaptive motivational profiles for different types of pupils based on the relations of each type of beliefs. Strong

agency beliefs of both ability and effort together with an emphasis on effort as a means for achieving goals have shown to indicate positive motivational orientation, whereas strong means-ends beliefs of ability have often been noted to relate to weaker agency beliefs, indicating a less favorable motivational profile (Little et al., 2002; Malmberg & Little, 2007; see also Lopez, 1999). The way action-control beliefs are manifested into action is also related to the way in which an individual perceives intelligence i.e. whether s/he believes abilities are fixed or malleable (Dweck, 1986; Malmberg et al., 2008).

2.1.2 Academic self-concept

The theoretical background for the current research on academic self-concept lies in the model of self-concept by Shavelson and his colleagues (1976). However, the roots of self-concept research dates back to the work of William James (1890) whose idea of the self as a multifaceted entity, which can be divided into knower-part (I) and knowing-part (me), has been an important inspirator of self-concept theory (James, 1890; Marsh et al., 2017). According to Shavelson and his colleagues' (1976) classical description, self-concept is a person's perception of themself, which develops through experiences with the environment. Self-concept is a considered to be a multifaceted and hierarchically organized construct through which general self-concept forms the apex of the hierarchy, which falls into academic and non-academic self-concept (Shavelson et al., 1976). Both non-academic and academic self-concepts can be divided into sub-dimensions. The non-academic self-concept falls into social, physical and emotional self-concepts, which in turn can be separated into more specific sub-dimensions. Academic self-concept can furthermore be divided into domain-specific self-concepts (e.g., mathematical or verbal self-concepts) and even more specifically to self-concepts that relate to specific school subjects (e.g., mathematics, languages, science, etc.) (Marsh & O'Mara, 2008).

Studies (Marsh et al., 2017, 2018; Marsh & Craven, 2006; Marsh & O'Mara, 2008) have shown that domain-specificity is an especially important feature of academic self-concept, i.e., different academic self-concepts have a strong correlation with achievement in the specific corresponding subject, whereas the correlation with other self-concepts can be near zero. This domain-specificity is visible in a way that academic self-concepts are measured. Academic self-concept measures are comprised of evaluative judgements about a person's behavior against self-set standards which arise from internal or external comparison processes (Parker et al., 2014). This evaluative aspect differentiates academic self-concept from self-efficacy, which shares many similarities with academic self-

concept but is still theoretically and empirically a different construct (Bong & Skaalvik, 2003; Marsh, Pekrun, et al., 2019; Parker et al., 2014).[4]

Academic self-concept has been a target of research because of its relation to academic achievement, but also because it has been seen as an important educational goal itself. Studies have shown that a strong academic self-concept can lead to many positive outcomes and success in life. Academic self-concept has been shown to be an important factor explaining educational choices such as course selections (Marsh & Yeung, 1997), university entry (Parker et al., 2012) and educational attainments (Guay et al., 2004). These effects can be very long-lasting. For example, in their longitudinal study, Guay et al., (2004) found that

[4] Bandura's (1977) concept of self-efficacy shares many similarities with both academic self-concept and Skinner's action-control beliefs (especially agency beliefs, see Lopez 1999; Walls & Little, 2005). Perceived competence is central to all these concepts, but concepts differ in the way it they are described. According to Pajares and Schunk (2001), self-concept researchers see perceived competence as individuals' perceived ability, where it indicates the skills and abilities that individuals believe they have, whereas self-efficacy researchers describe competence as individuals' confidence to succeed in the given situation. Therefore, agency and control expectancy beliefs can be seen as comprising elements of both of these, as they include both individuals' perceptions of their own abilities as well as expectations of success (Skinner et al., 1988, 1990). However, agency and control expectancy beliefs are usually described and measured at a more general level (see for example Skinner et al., 1988) than either of the other two constructs, which are described on a more domain-specific level. Yet, there is also a difference in the level of measurement between academic self-concept and self-efficacy (see Bong & Skaalvik, 2003). Academic self-concepts are often measured by subject-specific items, focusing on pupils' perceptions of their abilities in specific subject areas, whereas self-efficacy research utilize more task-specific items, which focus on pupils' conviction of their confidence to carry out the described task (Parker et al., 2014). In addition, even though all three concepts share many antecedents, there are also some essential differences in the mechanisms through which these concepts are expected to form (see Bong & Skaalvik 2003). Self-concept is most strongly influenced by context and social comparison (Marsh, Trautwein, et al., 2008), whereas previous mastery experiences are the main building blocks of self-efficacy (Bandura, 1977, 1993; for a review see Usher & Pajares, 2008). This is due to individuals evaluating their efficacy against certain goals, and prior experiences of success and failure are therefore the most influential reference points for individuals' efficacy expectations. However, also in the theory of self-efficacy, the role of the environment and social comparisons are seen as important, especially if the situation is new (Bandura, 1977, 1993). Again, action-control beliefs fall somewhere in the middle of these theories, resulting partly from previous experiences but also being influenced by the context (Geldhof & Little, 2011; Little et al., 2002). Studies in the field of action-control theory have detected that agency and control expectancy beliefs seem to be more influenced by the educational contexts than means-ends beliefs (Little et al., 2001, 1995; Little & Lopez, 1997; Malmberg & Little, 2007; but, see Liu & Yussen, 2005).

academic self-concept predicted educational attainment ten years later over and above prior achievement. However, maybe the most important aspect of academic self-concept is still its reciprocal and predictive relation to academic achievement (Guay et al., 2003; Marsh et al., 1988; Marsh & Craven, 2006; Marsh & Martin, 2011; Marsh & O'Mara, 2008).

2.2 Development of competence beliefs and thinking skills at school

Cognitive processes and individuals' awareness of these develop rapidly during childhood and adolescence (Demetriou et al., 2011). According to Demetriou and his colleagues (2011), self-awareness and self-evaluations of cognitive processes develop in cycles, where self-evaluations are accurate in each developmental stage, but become more inaccurate again before the next stage of development. Studies have indicated four important developmental cycles that occur at the ages of 0-2, 3-7, 8-12 and 13-18 (Demetriou et al., 2003, 2011; Demetriou & Kazi, 2006). According to Demetriou, Kyriakides and Avraamidou (2003), consciousness is the link between mind and personality and therefore developing consciousness of self and cognitive processes brings personality and emotional influences into cognitive functioning as they come to moderate these processes.

For the purpose of this study, studies I, III and IV focused on the timespan when children were ten to twelve years old, and study II took place when children were thirteen to fifteen years old. Thus, they were going through the last two developmental cycles presented in the theory espoused by Demetriou et al. (2003; 2011). As cognitive processes and self-awareness develop, children's understanding of action and themselves as agents sharpens. These maturation processes are reflected in action-control beliefs, which begin to differentiate as children grow (Little et al., 2002). Young children are not able to make a difference between the causes behind certain occurrences and before the age of approximately eight, children's thinking can be described as egocentric, as they have a tendency to overestimate their own role as agents (Geldhof & Little, 2011; Skinner et al., 1998). At the age of 9 or 10, children begin to be able to distinguish tasks according to the means they require, which means that they begin to understand the implications of succeeding at a task (Little et al., 2002; Little & Lopez, 1997; Malmberg et al., 2008; Roque et al., 2014; Skinner et al., 1998). Children begin to understand when success depends on luck or chance and when it is a matter of effort or skill (Little & Lopez, 1997; Skinner et al., 1998). At the same time, children's understanding of effort and ability becomes more differentiated and they begin to understand the interdependent nature of those constructs (Nicholls, 1978). In other words, children start to realize that showing greater effort might imply lesser ability (Nicholls, 1978). However, it has to be

noted that current research has challenged this view and suggested that in familiar and meaningful contexts, young children's competence judgements may be much more similar to older children's than it had been understood before (Cimpian, 2017; see also Eccles et al., 1993).

Longitudinal studies on the development of children's self-beliefs during school years have usually detected a declining trend in children's self-beliefs over time (Eccles et al., 1993; Frenzel et al., 2010; Jacobs et al., 2002; Pintrich & Schunk, 1996; Wigfield & Eccles, 2000; for a review, see Muenks et al., 2018; Stipek & MacIver, 1989). This finding is partly explained by the maturing and developing processes explained above, but also because the importance of social comparisons become more significant as children grow. However, as Geldhof and Little (2011) have pointed out, the developmental trajectories for specific beliefs can vary even though the mean level of self-beliefs would show a general decrease. For example, in Little, Stetsenko, and Maier's (1999) study, pupil's agency beliefs of ability increased from grade 2 to grade 11, while agency beliefs of effort decreased. Moreover, regarding academic self-concept, studies have shown that trajectories for different self-concepts can vary. Studies (e.g., Marsh et al., 2017) have suggested a curvilinear relation between age and self-concept where a decline in self-concept in preadolescence is followed by an increase in late adolescence (Jacobs et al., 2002; Marsh, 1989). Different competence beliefs can also develop and change at different paces (Geldhof & Little, 2011).

Self-beliefs develop in a reciprocal relationship with achievement (Eccles & Wigfield, 2002; Little et al., 1999; Marsh & Martin, 2011; Pajares & Schunk, 2001; Skinner et al., 1998). Therefore, experiences of success or failure shape individuals' expectancies and beliefs of their own competence, and vice versa, good achievement reinforces the trust in one's own academic abilities (Marsh & Craven, 2006; Skinner et al., 1998). School as an environment provides challenges and individuals can accomplish tasks and repeatedly get feedback on their own performance. This feedback shapes individuals' perceptions of their academic abilities and themselves as learners. However, to understand the feedback from the environment, there should be a reference point, i.e. frame of reference (Skaalvik & Skaalvik, 2002) to reflect them on. These frames can be either internal or external and self-beliefs take their shape through comparison processes against these frames (Marsh, 1984, 1986; Marsh, Parker, et al., 2019; Müller-Kalthoff et al., 2017).

Internal comparisons can be divided into temporal (Albert, 1977) and dimensional (Möller & Marsh, 2013) processes. Temporal comparisons refer to processes in which prior performance is used as a reference frame against which the present performance is evaluated (Albert, 1977). In dimensional comparisons, children compare their own accomplishments in various school subjects to their achievements in other subjects and this influences their perception of themselves in each domain-specific self-concept (Möller & Marsh, 2013). Studies have

suggested that internal and external comparisons might have a different role in the formation of self-beliefs (Skaalvik & Skaalvik, 2002). As temporal processes usually tend to indicate improvement over time and thus be satisfying, they might be more pronounced in situations in which individuals want to boost their self-beliefs, whereas external comparisons might be exploited when a more accurate picture of self is needed (Wilson & Ross, 2000). For the purpose of this study, the external comparison processes in which other people serve as a reference frame for individuals' self-appraisals, are more important and will be presented next.

2.2.1 The role of external comparison processes in the formation of self-beliefs

According to Festinger's (1954) social comparison theory, human beings have an inborn need to compare themselves to other people. This need drives individuals to social interaction with others and gives them an opportunity to evaluate their opinions and abilities. These comparisons are important building blocks for the self and without them individuals' perceptions of themselves would be more unstable and unrealistic, as they would lack an important reference frame. Studies on social comparisons (Ruble et al., 1980; for a review, see Dijkstra et al., 2008) have shown that comparison processes affect children from when they are small, but their purpose is different from those of older children (Stipek & MacIver, 1989). Small children rely on comparisons to acquire information about appropriate behavior, but the evaluative aspect of comparison processes becomes more and more important as children grow (Dijkstra et al., 2008).

People usually compare themselves to others who are similar to them (Festinger, 1954). For example, at school, children have a tendency to compare themselves with others who are of the same gender (Blanton et al., 1999; Thijs et al., 2010). However, similarity can also be many things other than gender. According to Festinger (1954), people usually compare themselves to others who they resemble in terms of abilities or opinions. If the group is very different from the individual, individuals have three options: either they can try to change themselves to be more like the group; they can try to change the group to be more similar to them; or, if the discrepancy between their own self and the group is too wide, they might leave the group. If leaving the group is not possible, Festinger points out that individuals can also redefine the comparison group and exclude some of the people from comparisons. In the context of school, this finding is interesting, as it could mean that if pupils think that they are very different from other classmates, they might place more emphasis on comparisons with only those who are similar to them.

However, it is important to note that in his theory, Festinger (1954) differed between the formation of opinions and the formation of abilities. In social comparison processes about abilities, Festinger states there is "a unidirectional

drive upward", as people have an inborn tendency to compete. Therefore, individuals seem inclined to compare themselves to other people who are slightly better. However, downward social comparisons are also possible (Dijkstra et al., 2008). Studies have suggested that on some occasions, individuals could rely on the information from downward comparisons, especially when they want to boost their self-confidence (Müller-Kalthoff et al., 2017).

Even though social comparison processes are mainly implicit and happen automatically and often unconsciously (Zell & Alicke, 2010), processes can vary in their nature. A distinction can be made between deliberate and forced comparisons (Huguet et al., 2009; see also, Dai & Rinn, 2008). Deliberate comparisons refer to self-initiated comparison choices, whereas forced comparisons (Diener & Fujita, 1997) refer to situations in which individuals are exposed to comparisons without a choice. These situations, such as a teacher reading exam grades aloud or returning tests to pupils with information about the distribution of grades, are common at school where children continuously get information about their own achievements in relation to others (Dijkstra et al., 2008; Marsh et al., 2004). Deliberate, upward comparisons, in which individuals have more initiative, can function as enhancive and motivating and thus lead to (positive) *assimilation effects* and even better achievements (Blanton et al., 1999; Huguet et al., 2001) (see Chapter 2.4.2). Then again, forced comparisons with high-achieving peers can influence the self negatively and lead to harmful effects (such as the *Big-Fish-Little-Pond effect BFLPE*, see Chapter 2.4.2) (Huguet et al., 2009; Seaton et al., 2008). Thus, different simultaneous processes can partly explain differing and even controversial findings in the research on peer effects (Müller-Kalthoff et al., 2017; Seaton et al., 2008; Skaalvik & Skaalvik, 2002).

Information for the formation of self-beliefs can be gathered from multiple sources (Skaalvik & Skaalvik, 2002). For example, in the context of school, pupils can compare themselves in relation to other pupils in the whole school, their class or then comparisons can be made in smaller peer groups. Different groups can have "a different realm of relevance" for individuals and their effect on the formation of self can differ (Festinger, 1954). Research in the field of local dominance theory argues that when multiple frames of references are available, people tend to rely on the information from the most proximal reference frame (Zell & Alicke, 2010). Zell and Alicke (2010) have introduced the core rationales behind local dominance hypothesis that is based on the idea of the importance that small groups have had for human beings throughout history. Small groups have been essential for human beings to survive and evolve. Small groups, such as family or friends, also usually capture many emotions and thus they are highly valuable for individuals. Thus, their meaning for the self-formation process can be more important than other groups (Festinger, 1954). In addition, human beings have been "habitually exposed" to comparisons in small groups, as children usually grow up in small families and only gradually start to interact with more

people (Zell & Alicke, 2010). Therefore, it is understandable that individuals tend to rely more on the information that they get from a small, local group than from larger units, even when they know that the information might be biased (Zell & Alicke, 2010). Even though local dominance hypothesis has been tested mainly in laboratory settings, studies proceeded in the school context have given some support to it, as class-level comparisons have shown to be more important for the formation of self-beliefs than comparisons at the school level (Marsh et al., 2014).

Because of external comparison processes, the same objective accomplishments can be interpreted differently in different contexts and therefore lead to differing self-beliefs regarding the composition of the reference group (Marsh & Hau, 2003). Therefore, the way pupils are allocated into groups and classes (see Chapter 2.3) can be influential and have an effect on pupils' learning.

2.2.2 Gender differences in the development of self-beliefs

It has been noted that both action-control beliefs and academic self-concept differ by gender. Regarding action-control beliefs, prior studies have found that girls generally stress more about the importance of effort as means of achievement than boys do, whereas boys trust their abilities and rely on luck (Niemivirta, 2000). A general tendency seems to be that, compared with boys, girls see themselves as more industrious, but underestimate their abilities more easily (Malmberg et al., 2008). Gender differences seem to be especially pronounced when the focus is on domain-specific beliefs (Diseth et al., 2014; Eccles et al., 1993; Stipek & Gralinski, 1991), whereas at a more general level, the differences are smaller (Niemivirta, 2000; Stetsenko et al., 2000). Domain-specific self-concepts differ in a stereotypic way so that boys usually have stronger self-concept domains in "masculine" subjects such as science or mathematics, whereas girls outperform boys in verbal subjects (Marsh, 1989; Skaalvik & Skaalvik, 2004).

Similar findings have been reported regarding other motivational constructs (Butler & Hasenfratz, 2017; Eccles et al., 1993; Meece et al., 2006). Overall, girls seem to adopt a more positive pattern of academic motivation characterized by high effort and good accomplishments, but also at the same time by a lack of self-confidence (Butler & Hasenfratz, 2017; Parker et al., 2018). Paradoxically, even though many studies have shown that girls outperform boys in many school subjects (Voyer & Voyer, 2014), boys constantly rate their abilities higher than girls do (Parker et al., 2018). Finland has been one of the example countries in OECD's Programme for International Student Assessements (PISA), in which gender differences in achievement have been favoring girls (Leino et al., 2019; OECD, 2020; Vettenranta et al., 2016). For example, in the most recent assessment (PISA 2018), Finnish girls outperformed boys in all three measured subjects (i.e. reading, mathematics and science) (Leino et al., 2019). The gender gap was especially pronounced in reading and science, in which the difference in

the achievement scores between girls and boys was noticeably larger than the OECD average (Leino et al., 2019; OECD, 2020). In reading, the gender difference favoring girls was especially pronounced among the highest- and lowest-performing pupils, whereas in mathematics the gender difference was less visible among the best performing pupils (Leino et al., 2019). This finding is in line with previous studies suggesting that the strength of gender differences may vary at different points of achievement distributions (Reilly et al., 2015).

Both girls and boys detect a decline in their competence beliefs during the school years (Jacobs et al., 2002), but the rate of decrease varies between genders. Overall, it seems that gender differences emerge at an early age (Butler & Hasenfratz, 2017). Studies have reported typical domain-specific gender differences from the beginning of the first grade (Eccles et al., 1993). Longitudinal studies have shown that developmental trajectories of different competence beliefs differ by gender domain specifically (Jacobs et al., 2002). In addition, gender differences in self-beliefs have been noted to be confounded with ability level, indicating that the change in girls' and boys' self-concepts may be different for the highest- and lowest-achieving pupils (Parker et al., 2018). Studies are inconsistent about the overall change in gender gap over time. For example, in the study on mathematics by Jacobs et al. (2002), the gender difference in competence beliefs favoring boys leveled off throughout the school years as boys encountered a faster decline in the mathematics-related competence beliefs than girls did. Then again, Frenzel et al. (2010) noted an increase in the gender gap as girls met a steeper decline in their mathematics interest than boys did over the years.

Gender differences in motivation have been explained in many ways (for a review see Hyde, 2014). Even though there is some evidence that biological predispositions explain some of the variance between genders, sociocultural context seems to have a stronger influence (Butler & Hasenfratz, 2017; Eccles, 2011). Therefore, the role of parents, peers and teachers is important in constructing and maintaining gender differences. As gender differences have been observed to emerge at such an early age, parental influence on producing gender stereotypic differences has shown to be important (Meece et al., 2006). For example, parents' expectations can be different for boys and for girls (Gunderson et al., 2012). When children go to school, they are exposed to a wider set of influences through peers and teachers. Studies have shown that children typically form peer groups with same-sex friends, which can intensify gender differences (Leaper, 2013; Thijs et al., 2010). As with parents, teachers' expectations of pupils have also been noted to differ by gender (e.g., see Butler & Hasenfratz, 2017; Gunderson et al., 2012). Teachers' role in the classroom is also important because teachers' impact on the learning environment in the classroom is essential. Studies have found that different contextual factors such as schooling culture can either attenuate or intensify gender differences. For example, Stetsenko et al. (2000)

found that in schooling contexts where gender stereotyping was stronger, gender differences were also more pronounced.

Even though the gender differences presented above are consistently found in research, it should be noted that typically the effects of gender are small and the variance between genders is markedly smaller than variance within gender (Butler & Hasenfratz, 2017; Hyde, 2005). Therefore, the gender similarity hypothesis proposed by Hyde (2005) emphasizes the similarity between genders and notes that gender differences should not be exaggerated.

2.3 Practices of tracking

The question of how to allocate pupils into schools and teaching groups is a universal question that every nation and educational system must answer. Even though it might seem like a simple, practical question, the amount of literature and research concentrating on it reveals that it is multi-faceted. Different allocation and grouping practices can be seen as an attempt to manage the diversity and individual differences in education and as a way to optimize teaching so that it would give maximum benefits for all pupils (Dupriez et al., 2008; Gottfried, 2014). In addition, the question related to grouping and tracking practices is a question concerning educational equality (Hanushek & Wößmann, 2006; Hattie, 2002). Teaching all children in the same comprehensive schools or segregating them into different tracks either within or between schools is a choice that can either promote or hinder educational equality.

Different allocation and grouping practices are studied under different concepts such as *tracking*, *streaming* and *ability-grouping* and the use of these seem to be somewhat inconsistent (e.g., Betts, 2011; Boaler et al., 2000; Chiu et al., 2017; Saleh et al., 2005; Trautwein et al., 2005, 2006). For example, Betts (2011) makes a difference between *tracking* and *streaming* and states that they refer to practices where pupils are divided into schools (*streaming*) or classrooms (*tracking*) by their abilities or achievement. Similarly, Chiu (2017) uses the term *tracking* when referring to groupings inside classrooms, but uses *streaming* to refer to groupings within schools, i.e., at the grade level. In addition, the terms *between-school tracking* and *within-school tracking* have been used to describe tracking practices at different levels (e.g., Van Houtte & Stevens, 2015). *Ability grouping* seems to be used mainly to refer to practices inside schools (e.g., Boaler et al., 2000; Ireson et al., 2001; Saleh et al., 2005), but again definitions vary and include many different types of grouping practices (for a review, see Steenbergen-Hu et al., 2016; see also Hattie, 2009).

In this study, the term *tracking* represents a wide meaning including different allocation and grouping practices. More specifically, to describe the Finnish tracking system (see more in Chapter 2.5), the term *opt-in-tracking* suggested by Trautwein and colleagues (2005) has been used. Trautwein and colleagues (2005)

have created a classification scheme, which enables comparisons between tracking practices. Accordingly, different practices can be classified based on three categories. Firstly, the institutional level where tracking takes place should be considered (i.e., is the tracking occurring at or within the school level). Secondly, the focus should be on the role of achievement in determining the track placement. Achievement can be the only factor diving pupils into tracks, but tracks can also be based on other aspects such as pupils' interests or parental educational goals. If prior achievement is the main factor determining the track, then Trautwein et al. (2005) suggest that the term achievement grouping is the best term to use, whereas the term opt-in tracking is suggested for use in situations in which other factors also influence the placement. In the Finnish tracking system, pupils can apply to receive emphasized teaching that is typically organized in selective classes, which emphasize some selected subjects and vary in acceptance criteria (see more in Chapter 2.5). The third aspect, which Trautwein and colleagues (2005) propose to consider, is the impact of tracking on determining individuals' future careers. In other words, is the placement in a certain track limiting pupils' future educational choices or do all possibilities remain open regardless of pupils' track (see also Trautwein et al., 2006).

Arguments favoring tracking highlight the importance of individuality in education, and claim that sorting pupils into homogenous groups based on, for example, their abilities make it possible to customize teaching so that it would better serve the range of needs of pupils and result in benefits for all (Tieso, 2003). However, arguments against tracking claim that it produces educational inequality (Hanushek & Wößmann, 2006; Schütz et al., 2008) as the gains from tracking may not be distributed evenly across all pupils (Hattie, 2002).

Practices of tracking differ greatly between countries and education systems (Dupriez et al., 2008; OECD, 2016). There are countries like Austria or Germany that track pupils into "differing-ability schools" from the age of 10 (Hanushek & Wößmann, 2006), but the most common age for between-school selection according to OECD (2016) seems to be 15, meaning that pupils are tracked into different schools only after lower secondary school. However, before that, many countries used forms of tracking inside schools that varied according to when tracking occurs (e.g., tracking during class allocation) or how it occurs (e.g., within-class achievement grouping or course-level ability grouping) (Trautwein et al., 2005). Countries and education systems also vary regarding how explicit or implicit the tracking is (Trautwein et al., 2006). In some countries, tracking is highly visible and can even be prescribed in law, while in other countries, it can be so invisible that not even the children know that they have been tracked.

2.4 Effects of tracking

Effects of *tracking*, *streaming* or *ability grouping* have been studied widely in several fields of education ranging from the economics of education to educational psychology (Betts, 2011; Burns & Mason, 2002; Chiu et al., 2017; Collins & Gan, 2013; De Fraine et al., 2003; Duflo et al., 2011; Hattie, 2002; Kang, 2007; Kindermann, 2007, 2016). Collins and Gan (2013) have proposed that studies could be divided roughly into those that focus on *direct tracking effects* and those focusing on *peer effects* (see also Pallas et al., 1994). Accordingly, *tracking effects* refer to different instructional practices such as teaching methods or quality of teaching, whereas *peer effects* refer more to compositional effects resulting from changes in group compositions. Therefore, the terms *contextual* or *compositional effect*s are also often used when referring to *peer effects* (e.g., Burns & Mason, 2002; Dicke et al., 2018; Harker & Tymms, 2004; Willms, 1985). In this study, the term *tracking effect* is used in a wide sense to refer to all consequences that different grouping practices can have either on individuals' learning or educational equality, whereas the term *peer effect* is used when referring to a more specific tracking effect at the classroom level.

As practices of tracking differ widely between and within nations and school systems, effects of tracking have also been analyzed at different levels (e.g., at school and classroom levels and even at course level) (e.g., Marks, 2010; Burke & Sass, 2013; Gottfried, 2014).[5] Therefore, resulting from different research traditions, methods and research settings, the findings of prior research are quite incoherent and there seems to be relatively little consensus over the question of how tracking influences an individual's learning. Next, I will review some findings of prior studies on tracking effects. For this study, the focus will mostly be on studies that have proceeded at the class level. First, I will go through the literature of peer effects on academic achievement and self-beliefs and then I will briefly summarize the findings of prior studies from the perspective of educational equality.

[5] In addition, peer effects have also been studied extensively in smaller peer or friendship groups (Altermatt & Pomerantz, 2005; Cooc & Kim, 2017; Ryan, 2000; Wentzel, 2017; Wentzel & Caldwell, 1997). These studies have analyzed the influence of different types of peer group (groups of friends, proximity groups, or frequent "hangout" groups) on children's behavior, school achievement or motivation (Kindermann, 2016). However, for the purpose of this study, the focus has been on studies conducted at the classroom level and therefore research on small peer groups have not been the focus here.

2.4.1 Effects on achievement: Peer Spillover effect

Most of the studies analyzing peer effects in education have focused on the effects on academic achievement (e.g., Burke & Sass, 2013; Chiu et al., 2017; Collins & Gan, 2013; Gottfried, 2014). These studies have analyzed the hypothesis sometimes referred to as the Peer Spillover effect (e.g., Burke & Sass, 2013; Dicke et al., 2018; Duflo et al., 2011), which suggests that studying in a well-achieving peer environment has beneficial spillover effects on achievement.

As different grouping practices change the composition of learning groups and make them more homogeneous than they would otherwise be, many studies (e.g., Collins & Gan, 2013; Duru-Bellat & Mingat, 1998) have analyzed whether studying in a homogeneous (i.e. tracked by ability) versus heterogeneous (non-tracked) learning environment is beneficial for all individuals. Studies have suggested that peer effects might differ across pupils depending on their ability level (Burke & Sass, 2013; Carman & Zhang, 2012; Gottfried, 2014; Kang, 2007) or other individual attributes such as gender or ethnic background (Hoxby, 2000). For example, in an experimental study by Saleh and his colleagues (2005), authors found that grouping by ability most benefitted the average-achieving pupils, whereas high-ability pupils accomplished equally well both in grouped or non-grouped learning environments. Kang (2007) showed that pupils' own achievement level influences the interaction they have in classrooms. Accordingly, weak pupils interact more with other weak pupils and controversy, strong pupils associate mostly with their well performing classmates. Therefore, if pupils are grouped according to their abilities, weak pupils might suffer the most. This finding, which has been the main argument for tracking opponents, has been confirmed in many studies and therefore it has been stated that low-achieving pupils would benefit from mixed grouping practices, because they would have a chance to benefit from the presence of better performing pupils (Duru-Bellat & Mingat, 1998; Saleh et al., 2005).

However, some studies claimed that tracking would be beneficial for all pupils (Collins & Gan, 2013; Tieso, 2003). For example, Collins & Gan (2013) found out that sorting pupils into homogeneous groups significantly improved all pupils' mathematics and reading scores. Similarly, Duflo & Dupas Kremer (2011) stated that tracking pupils by prior achievement raised scores for all pupils, even those assigned to lower-achieving tracks. Indeed, some studies have even suggested that low-achieving pupils might suffer from the presence of better peers (Carman & Zhang, 2012) and therefore would benefit from a homogeneous learning environment.

In addition, studies have suggested that peer effects on achievement might differ depending on the academic subject (Carman & Zhang, 2012; Gottfried, 2014), but results have not been systematical in terms of different subjects. Studies have also claimed that the strength of peer effects might be most pronounced at the primary school level where children spend more time in the same classroom

than in middle or high school (Burke & Sass, 2013). However, this view has also been confronted as studies on the Big-Fish-Little-Pond Effect (see Chapter 2.4.2) have systematically found that the size of this peer effect is larger for older pupils (Fang et al., 2018; Marsh et al., 2015).

Another branch of studies (e.g., Burns & Mason, 2002; De Fraine et al., 2003; Dicke et al., 2018; Hienonen et al., 2018; Peetsma et al., 2006) has examined tracking effects as compositional effects and examined how the composition of the classroom (or school) affects individuals' learning outcomes. Therefore, they have examined how the proportion of various student groups in a classroom affect individuals' learning outcomes. These groups include low versus well-achieving pupils (De Fraine et al., 2003; Peetsma et al., 2006), high SES pupils (Belfi et al., 2016; for a review of peer socioeconomic status on achievement see van Ewijk & Sleegers, 2010), girls versus boys (De Fraine et al., 2003), special need pupils (Hienonen et al., 2018; Hienonen, 2020), and pupils from different ethnic backgrounds (Hornstra et al., 2015; Peetsma et al., 2006). For example, De Fraine and her colleagues (2003) analyzed the changes in children's language achievement and examined whether there were differences in children's language achievement between different types of schools and classes. Their findings indicate that group composition would be an important factor determining the development of children's language achievement, as they found that it developed more rapidly in classes where the average ability level was higher and where there was a large proportion of girls. Similarly, Hienonen (2020; 2018) concluded in her doctoral thesis that class composition does matter. Hienonen examined the composition of classes from the perspective of special educational needs (SEN) pupils and revealed that group composition played an important role in the learning results of the pupils with and without SEN. Pupils with SEN seemed to benefit from the presence of other SEN pupils in the classroom while pupils without SEN seemed to be distracted by the presence of too many SEN pupils in the same class. Their performance level was lower compared to pupils in a class with a smaller proportion of SEN pupils. Also, van Ewijk & Sleegers (2010) conclude in their meta-analyses of compositional effects of peers' socioeconomic status that peers matter, but they also highlighted the complexity of investigating compositional effects and argue that results in different studies vary because of inaccurate measures and differing methods.

This heterogeneity of tracking and peer effect studies makes it difficult to draw any strong conclusions about the influence that tracking may have on individuals' achievement. Conclusions in meta-analyses and reviews have also been inconsistent. For instance, in their second-order meta-analyses of ability grouping, Steenbergen-Hu, Makel, & Olszewski-Kubilius (2016) explored the effects of various forms of ability grouping on pupils' academic achievement and concluded that pupils of all ability levels seem to benefit most from ability grouping. More specifically, they found that all other forms of ability grouping (i.e., within-class

grouping, cross-grade subject grouping and special grouping for giftedness) seemed to be beneficial, except between-class grouping (i.e., a practice by which pupils of the same grade are assigned to different level classes based on their prior achievements). However, based on his meta-analyses, Hattie (2002; 2009) concluded that the average effects of tracking or other grouping practices on learning outcomes would be minimal, if they existed at all. In his conclusions, he proposes that the most important things for learning seem to happen inside the school, at the classroom level, and stresses the importance of teachers, but also the influence of peers.

Studies (Dicke et al., 2018; Harker & Tymms, 2004; Marks, 2010; Nash, 2003) have also questioned positive peer effects on achievement altogether. As an attempt to solve the theoretical paradox between findings of Big-Fish-Little-Pond effect research (i.e. negative peer effects on academic self-concept) (Marsh, 1987; Marsh & Parker, 1984; Marsh et al., 2000) and studies suggesting positive peer spillover effects on achievement, Dicke and her colleagues (2018) conducted a large longitudinal study with improved methodology to analyze the school-level compositional effects. Their results contest findings of positive peer effects and claim them to be "phantom effects" that disappear when analyses are performed with appropriate statistical controls for measurement error and pre-existing differences. Therefore, the results from Dicke and her colleagues (2018) also emphasize the importance of adequate statistical methods and research settings when the complex dynamics of school and class compositional effects are explored (see also Televantou et al., 2015).

2.4.2 Effects on self-beliefs: Big-Fish-Little-Pond & Reflected Glory

Peer effects on self-beliefs have been analyzed most coherently in the field of academic self-concept, where a substantial amount of research has identified two possible mechanisms through which peers influence an individual's academic self-concept: Big-Fish-Little-Pond and Reflected Glory effect (Marsh et al., 2000; Marsh, Seaton, et al., 2008; Marsh, Trautwein, et al., 2008; Trautwein et al., 2006, 2009). Both hypotheses test the effects of overall ability level in the school/classroom on an individual's academic self-concept. Therefore, studies on the Big-Fish-Little-Pond and the Reflected Glory effect belong to the same branch of studies analyzing compositional peer effects (see below in Chapter 2.4.3).

Research on the Big-Fish-Little-Pond effect (BFLPE) (Marsh, 1987; Marsh & Parker, 1984) is grounded on the assumption that pupils' academic self-concept is influenced by comparisons with their peer groups' performance level. According to the classical formulation of BFLPE, the average ability level of the school will predict an individual's academic self-concept negatively, even though at an individual level the relation between academic self-concept and achievement is

positive (Marsh, 1987; Marsh & Parker, 1984). Therefore, the theory predicts that equally able pupils have lower academic self-concept in a group in which the average ability level is high than in a group where the average ability level is low. In other words, it predicts that a child studying in a high-achieving group will develop a more negative academic self-concept than s/he would in a less proficient group.

The core hypothesis of the BFLPE was formulated already in 1980s and since then it has been confirmed in several studies (e.g., Fang et al., 2018; Marsh et al., 2004, 2014, 2015; Marsh, Parker, et al., 2019; Marsh, Seaton, et al., 2008; Marsh & Hau, 2003). Extensive research over the years has shown that BFLPE is a viable theory in explaining how a peer group's achievement level influences an individual's academic self-concept (for reviews of the research over the years see, Marsh, Seaton, et al., 2008; Marsh & Seaton, 2015). Research has shown that BFLPE is a phenomenon that can be widely generalized to many countries and cultures (Marsh et al., 2015; Marsh, Parker, et al., 2019; Seaton et al., 2009) as well as to different subject domains (Guo et al., 2018). BFLPE was initially formulated at the school level (Marsh, 1987; Marsh & Parker, 1984). Therefore, the early studies examined the effect of average achievement level of the school on individual pupil's self-concept. Later studies have expanded the analyses of BFLPE and showed it to be tenable also at the class level (Guo et al., 2018; Marsh et al., 2015; Marsh & Seaton, 2015). Indeed, in a study by Marsh, Kuyper and their colleagues (2014) in which BFLPE was analyzed simultaneously at the class and school level with a three-level model, the results supported the local dominance theory (Zell & Alicke, 2010, see also Chapter 2.2.1) and showed that BFLPE was more pronounced at the class level than at the school level .

Studies have also shown that the effects of BFLPE can be long lasting (Marsh et al., 2007) and that both bright pupils and the weakest ones suffer from BFLPE similarly (Marsh & Hau, 2003). Even though it would be reasonable to expect that high-achieving pupils might not suffer from BFLPE as much as their weaker classmates and some studies have found small support for this argument (Coleman & Fults, 1985; see also, Marsh & Seaton, 2015), more support has gained the argument that BFLPE strikes all pupils similarly despite ability level (Marsh et al., 2007, 2014; Marsh, Seaton, et al., 2008). BFLPE has also been found to be reasonably similar for both girls and for boys (Loyalka et al., 2018; Marsh et al., 2007), even though academic self-concept has been noted to differ between genders (e.g., Marsh, 1989; Thijs et al., 2010; see also Chapter 2.2.2). According to Marsh and Seaton's (2015) review of over 30 years' research on BFLPE, most of the studies on BFLPE have examined the phenomenon at the lower secondary level. But a growing body of research on younger pupils have already provided evidence of the existence of BFLPE in primary school (Dicke et al., 2018; Guo et al., 2018; Marsh et al., 2015). However, BFLPE have shown to be stronger for older pupils (Marsh et al., 2015; Parker et al., 2019). Studies have suggested that

this could be the result of two factors. First, from age-based differences in ability stratification of school systems (i.e., stratification increases at upper levels of education) (Marsh et al., 2015; Parker et al., 2019) and secondly, from maturing cognitive processes that strengthen the role of social comparisons in the formation of self-concepts (Dicke, 2018; see also Chapter 2.2.1).

In addition to theoretical extensions, BFLPE research has also expanded in methodological approaches as advanced statistical methods have enabled more precise, complex and coherent analysis than those in the early years. Current research in the field of BFLPE is grounded on multilevel models, which makes it possible to take into account the hierarchical structure of the school-related data. Nowadays, multilevel models are considered a minimal condition in order to analyze BFLPE (Marsh, Seaton, et al., 2008; Marsh & Seaton, 2015).

In addition to BFLPE, studies have proposed another, competing mechanism, "Reflected Glory Effect", (RGE) which assumes that in addition to comparisons with peers, pupils' self-beliefs are affected by their perception of the ranking of their class compared to other classes (Marsh et al., 2000; Trautwein et al., 2005, 2009). In other words, children who study in selective classes or schools might get a boost to their self-beliefs, as they know that their class or school is highly ranked and difficult to get into. Despite studies having shown BFLPE to be more pronounced than the Reflected Glory effect (Marsh et al., 2000), it has also been shown that these competing effects can coexist and occur at the same time and therefore balance each other out (Marsh, Seaton, et al., 2008). This finding is in line with the theoretical assumption of multiple reference frames affecting the formation of self-beliefs (Skaalvik & Skaalvik, 2002, see also Chapter 2.2).

The implications of BFLPE and RGE research on the debate over tracking practices can be considered from two perspectives. Firstly, they show that tracking pupils into selective high-ability groups or tracks can have detrimental effects on their academic self-concepts and that even though these negative effects can be balanced with the glory effect stemming from "being better than the others" (Marsh et al., 2000), the overall negative effects are persistent. Therefore, these findings would speak against tracking. However, from the perspective of low-achievers, BFLPE phenomena could be interpreted to point out the opposite and defend practices where lower-achieving pupils would be allocated into their own teaching groups in order to avoid comparisons with high-achievers (Dupriez et al., 2008; Salchegger, 2016). Then again, children allocated to low tracks might suffer from an opposite Reflected Glory effect, as track allocation might shake their perceptions of themselves (Van Houtte, 2016; Francis et al., 2017).

2.4.3 Educational equality and tracking

As stated, the decision of whether or not to track or sort pupils can have implications not only for pupils' learning outcomes, but also for educational equality (Brunello & Checchi, 2007; Hanushek & Wößmann, 2006; Hattie, 2002). Many studies analyzing the equality effects of tracking have compared the overall achievement level and the distribution of achievements of different school systems and explored whether tracking leads to better results overall and, on the other hand, whether it influences the equality of education (for a review see Wößmann, 2009). Educational equality has been operationalized differently in different studies, for example Hanushek and Wößmann (2006) examined the deviation of test scores in tracked and non-tracked countries. In their analyses, they examined whether the difference between test scores from primary to lower secondary school are different in countries where tracking occurs at the lower secondary school level compared to countries where tracking occurs later. As a conclusion, they propose that early tracking seems to increase educational inequality as the deviation of test scores increased in countries with tracked school systems, whereas in non-tracked countries the deviation of scores decreased from primary to lower secondary school. In addition, their results suggested that tracking might have negative effects on the overall performance level of schools, even though this effect was not as strong as the effect on equality. Similarly, in their comparative study, Dupriez, Dumay and Vause (2008) concluded that social inequality measured by deviation in test scores was highest in countries that direct pupils into separate tracks already from the end of primary school. However, their findings of achievement differences between differently organized tracking systems pointed out the difficulty of system-level comparisons and showed that the effects of tracking on achievement may be determined by more detailed practices. Accordingly, the effectiveness of the school system was determined by grade retention practices, not by the stratification of the school system per se.

One way to examine the equality effects of tracking has been to focus on the impact of the family's socioeconomic status (SES) on pupils' performance in different tracking school systems (Maaz et al., 2008; Wößmann, 2009). Studies (Horn, 2009; Schütz et al., 2008) have shown that in early tracking countries, the relationship between pupils' performance and their family background is stronger than in countries with less stratified tracking systems, indicating that equality of educational opportunities is smaller in tracked systems. This has been explained with findings showing that track placements (Epple et al., 2000; Kelly, 2001) and school choices (Ball et al., 1995; 1996; Lauder, 1999; Vincent, 2001; Reay & Lucey, 2003) are strongly dependent on pupils' family background. Pupils from high SES families have better resources for making beneficial choices in the educational market (Ball & Vincent, 1998) and therefore pupils from high SES families are more likely to end up in high-ability tracks or schools where they

might experience achievement gains resulting from *direct tracking effe*cts (i.e. instructions, advanced curriculums, etc.) or *peer effects* (see Chapter 2.4).

2.5 Finnish classes with a special emphasis as implicit tracks

Classes with a special emphasis were introduced to the Finnish basic education system in the 1990s with reforms in educational legislation. Along the new basic education act, which came into force at the beginning of 1999, the schools had a chance to specialize by offering teaching with a special emphasis and the right to select pupils with aptitude tests (Kalalahti et al., 2016; Seppänen, 2006; Seppänen, Kalalahti, et al., 2015).[6] In addition, changes in the educational legislation changed the school place allocation policy and gave municipalities more freedom to organize it (Seppänen, 2003, 2006). For families, it offered a chance to participate in the allocation process more actively and opened up a chance for school choice (Varjo et al., 2014).

Changes in educational legislation were part of a larger societal wave of reforms, in which decentralization and deregulation of the state-centered governance were key words (Heiskala & Luhtakallio, 2006). Ideologically, changes in educational policy can be connected to global policy trends, in which neo-liberal market driven values became valued also in the field of education (Kalalahti et al., 2016). Traditional comprehensive school values emphasizing universalism and equal educational opportunities for all were challenged by neo-liberal values of competition and individualism (Rinne, 2000; Ylönen, 2009). The equality of educational opportunities has been a central value in Finnish education policy and a pervasive idea behind the formation of the Finnish comprehensive school system in the 1970s (Kalalahti & Varjo, 2012). The aim has been to give all children similar opportunities for education notwithstanding their gender, social background or residence (Ahonen, 2001). This emphasis on equality is strong in Finland and becomes visible in a way, how educational changes, for example school choice was carried out in Finland (Seppänen, Carrasco, et al., 2015; Varjo & Kalalahti, 2015). Even though parents can present a preference concerning their children's school place, this preference is only used in addition to other criteria. Therefore, the actual opportunities to choose are very restricted in Finland when compared internationally (Seppänen, 2003, 2006; Seppänen, Carrasco, et al., 2015). Choices are made mainly within a public school system, and families are not offered any official information on the differences between schools, as Finland does not support national testing of schools at the basic

[6] Even before that, municipalities had a chance to give special permission to some schools to establish special classes with emphasis on some subjects (Seppänen, 2003). Therefore, there were only a few classes (mainly music or languages) with a special emphasis before the educational reforms in the 1990s (Seppänen & Rinne, 2015).

education level (Seppänen, 2003, 2006; Seppänen, Kalalahti, et al., 2015; Varjo & Kalalahti, 2019; Wallenius, 2020). Therefore, the role of classes with a special emphasis is essential in the Finnish school choice model, as they offer an actual chance to select a school (Kalalahti et al., 2015; Varjo et al., 2014; Kosunen et al., 2020).

Since their establishment, classes with a special emphasis have become popular, especially in big cities, which offers options and choice. Options range from academic subjects (e.g., mathematics, science, languages) to arts (e.g., music, visual arts, sports). Studies have evaluated that in some cities approximately 30% of pupils study in a class with a special emphasis (Seppänen, Kalalahti, et al., 2015). Choosing a class with a special emphasis is usually possible either at the beginning of comprehensive school at the age of seven, before the 3rd or 7th grade (see Figure 2). Pupils are selected via aptitude tests, and the criteria for selection vary regarding the subject emphasized (Varjo & Kalalahti, 2015, 2019). Even though in practice classes with a special emphasis have only 1-2 hours per week more teaching in a selected subject, teaching is usually arranged so that children study all the time in the same selected group (Simola et al., 2015). Therefore, classes with a special emphasis have been seen to act as implicit tracks inside the comprehensive school system (Berisha & Seppänen, 2017; Varjo & Kalalahti, 2019). In addition, research has shown that classes with a special emphasis attract relatively high-achieving children from highly-educated families (Kalalahti et al., 2015; Kosunen, 2014; Kosunen & Seppänen, 2015; Seppänen, 2006; Seppänen, Kalalahti, et al., 2015). Thus, the equity effects of attending classes with a special emphasis have been debated (Kosunen, Bernelius, et al., 2016; Seppänen et al., 2012; Seppänen, Kalalahti, et al., 2015). Researchers have been worried about the "cream skimming" effect of these classes and argued that they endanger the equality of educational opportunities as they differentiate school paths (Berisha & Seppänen, 2017; Kosunen et al., 2020). It has been shown that classes with a special emphasis can increase within-school differences in learning results (i.e. differences between classes) (Bernelius, 2013; Kosunen, Seppänen, et al., 2016; Kuosmanen, 2020), which have already been large in Finland, compared to other Nordic countries (Yang Hansen et al., 2014).

Yet, selectivity based on pupils' own interest has been considered to be important by the policy promoters and other educational actors in ensuring the individuality of teaching (see in Seppänen, 2006; Varjo & Kalalahti, 2015). Especially parents have been keen to hold onto the option to choose a selective class (Kosunen & Carrasco, 2016). Studies have shown that parents believe that studying in a selective peer group would be advantageous for their child (Kosunen, 2014). Parents believe that the atmosphere in these classes would enhance pupils' motivation and enthusiasm for school work and thus be of benefit

to the child (Kosunen & Carrasco, 2016; Kosunen & Seppänen, 2015). However, before this dissertation, no empirical studies have examined these assumptions.

As the selection into classes with a special emphasis is not based on pupils' prior achievement, but on other aptitudes (Varjo et al., 2014; Varjo & Kalalahti, 2019), the term opt-in tracking within schools (Trautwein et al., 2005), might best describe the Finnish system. Selection criteria for emphasized subjects vary according to the emphasized subject and between municipalities (Varjo & Kalalahti, 2015). Usually schools arrange aptitude tests at which pupils can show their talent and interest in an emphasized subject. However, it must be noted that even though classes with a special emphasis can be described as an implicit tracking system within comprehensive school, all children still study within the same national core curriculum and get the same educational opportunities for upper secondary education (Seppänen, 2003). Actual choice between general or vocational upper secondary track is made only after basic education, when children are 16 years old (see Figure 2).

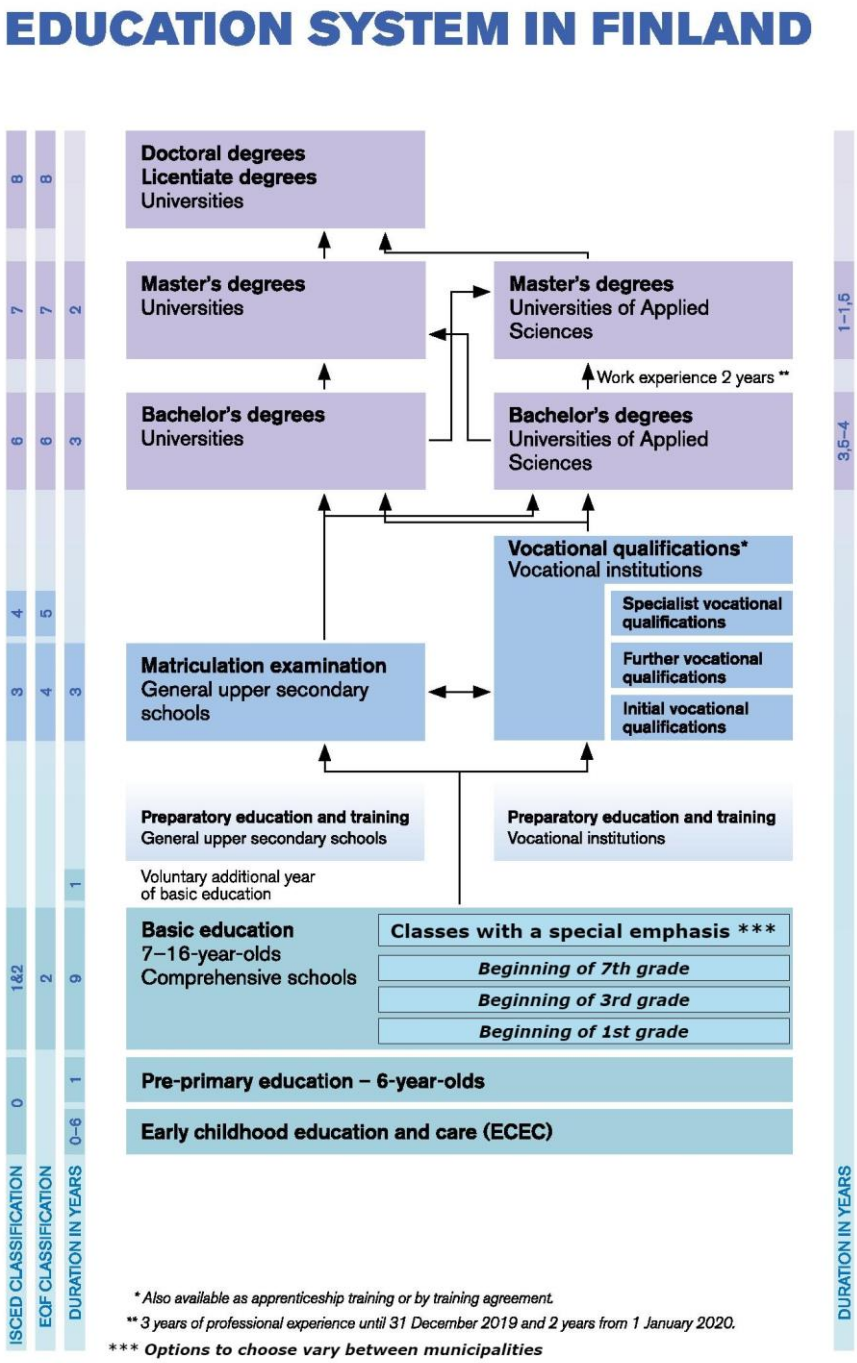

Figure 2 Structure of the Finnish education system (Ministry of Education and Culture, 2020)

2.6 Summary of the key concepts and contexts of the research design

The theoretical framework of this thesis is based on socio-cognitive theories of human development and learning (Bandura 1999, Vygotsky, 1978), which acknowledge the influence of social environment in the development of individual. The concepts and the research design of this study are summarized in Figure 3. This summary has been inspired by Bronfenbrenner's (1977; 2005) theory of human development, which has been used as an illustrative framework tying together the phenomena which have been described in the previous chapters. The central idea in Bronfenbrenner's (1977; 1979; 2001) ecological (later bioecological) system theory of human development is to examine an individual's development in its context. In his theory, Bronfenbrenner (1977) defined different layered ecologies or environments, which affect individual's development. The theory highlights the interrelated nature of these layers (systemic levels), emphasizing the idea that human beings create their environments, which again shape and influence their development (Bronfenbrenner, 2005).

The idea behind this thesis has been to explore the commonly held assumption regarding the beneficial motivational effects of classes with a special emphasis (see Chapter 2.5). Using Bronfenbrenner's (1977; 2005) theory as an illustrative frame enables to set this research in its context. Even though the focus has been on analyzing the development of competence beliefs and mathematical thinking skills of an individual child (see Figure 3), the analyses have intersected on the surface of the individual and microlevel (see Chapter 3.4).

In Bronfenbrennes (1977; 2005) theory, the microlevel is the closest level to the individual and contains all the environments in which the individual is self actively involved. An individual's immediate communities and surroundings such as family, peers and school, and inside school, the school class, belong to the microlevel (Bronfenbrenner, 1977; 2005). The microlevel is followed by the mesolevel, which refers to the relationships that an individual's microsystems have with each other (Bronfenbrenner, 1977; 2005). Thus, for example, the relationship between family and school (i.e., home-school collaboration) as well as the interaction between a child's own family and other families, formulate this level. Two outermost circles, exo- and macro level, differentiate by nature from micro and mesolevels, as in those, the individual does not function directly but rather they formulate the wider social structure, which influence to the nature of interaction within all other levels (Bronfenbrenner, 1977; Paquette & Ryan, 2001). The macro level comprises for example values, ideologies and laws in the society whereas the exo level refers to the level in which these upper level ideas are turned into action in forms of local policy practices etc. (Paquette & Ryan, 2001).[7]

[7] Later Bronfenbrenner (1992) expanded his theory by making the element of time more explicit and inserting the idea of *chronosystem* into it. *Chronosystem* refers to the elements of time and change,

Even though empirical analyses conducted in the articles which form part of this dissertation have not yielded results beyond individual and microlevels, it is good to acknowledge the influence of upper systemic levels. Using Bronfenbrenner's (1977; 2005) model enables us to see how macro level ideological and political changes during 1990s school reforms, led to changes in the exo level i.e., the local school choice policies and municipal practices and thus changed the local municipal school choice spaces (Varjo & Kalalahti, 2019) by giving families a chance to participate in the school choice of their children and by encouraging schools to specialize (see Chapter 2.5). [8] Together these upper level elements influence the compositions of schools and classes and thereby shape the realities of children's microlevel interaction by influencing to the compositions of microlevel groups.

In this thesis, focus has been on an individual's competence beliefs, which are central in many theories of motivation (e.g., Eccles & Wigfield, 2002). Therefore, in Figure 3, competence beliefs have been described as an ellipse, which forms one piece of the bigger ellipse of motivation. The ellipse of competence beliefs is closely connected to the ellipse of "cognitive functioning" as perceived competence is closely related to cognitive functioning and abilities (e.g., Bandura, 2004; Skinner et al., 1998). Even though competence beliefs can be seen as one piece of the motivation, theoretical approaches vary in terms of how closely they are related to other motivational aspects. Therefore, the ellipse of competence beliefs is not totally overlapped by the ellipse of motivation. [9]

In this thesis, motivation has been defined broadly, entailing many different theories and approaches (see Chapter 2.1). Motivation is thought to develop in interaction with the environment (Eccles et al., 1998; Urdan & Schoenfelder, 2006). The environment provides challenges and stimulus for an individual´s actions and strivings and a scene of actions for an individual to fulfill their inner

which permeable all levels (Bronfenbrenner, 1992; Paquette & Ryan, 2001) and could therefore be considered as the outermost circle in the model.

[8] Families' school choices are also influenced by other upper-level structures, such as social class. Class structure can be seen as a macro level structure, which molds the social realities of individuals' and thereby guide and constrain their actions (Bourdieu, 1989). Social class consists of capitals (social, cultural and economic) and dispositions, which has an influence on the families' school choice strategies (Ball et al., 1996; Bourdieu, 1989; Reay & Lucey, 2003). For example, Ball & Vincent (1998) investigated "hot" and "cold" knowledge related to school choices and showed, that middle-class families had more advantageous social networks (e.g. with other families), which gave them access to informal information related to school choice practices and therefore gave them benefits in school markets.

[9] For example, regarding the theories, which have been used in this study (i.e. action-control theory and theory of academic self-concept), action-control theory and other motivational theories (e.g., self-determination theory by Deci & Ryan 2000) can be thought to share more foundational antecedents whereas the roots of academic self-concept research are more linked into the research of self (see Chapter 2.1.2). Therefore, it can be thought that in the figure, academic self-concept could be posited a little bit further from from the ellipse of motivation than action-control theory, which essentially belongs inside the ellipse of motivation.

biological and psychological needs (Chang et al., 2017). In the context of school, children are constantly confronted with achievement situations which persuade them to act. The way an individual acts in these situations depends on situational factors (e.g., vitality level, working conditions at class etc.) and individual factors (e.g., personality, temperament) as well as an individual's prior emotionally embedded experiences and interpretations of those (e.g., did I succeed in this kind of task before and why) (e.g., Boekaerts, 1992; Chang et al., 2017; Little, 2002; Niemivirta, 2000; Pekrun, 2017; Rawlings et al., 2020). When these achievement situations and experiences happen repeatedly, an individual's perceptions of him/herself as an agent as well as perceptions of certain actions and goals evolve (Little, 2002). The development of these perceptions is closely connected to general cognitive development (e.g., reasoning skills, etc.; e.g., Demetriou et al., 2011), which influence the way an individual perceives him/herself and the environment (see Chapter 2.2).

In addition, the environment gives feedback and poses reference frames for an individual to evaluate his/her actions (Chang et al., 2017). For the purpose of this thesis, the social comparison theory by Festinger (1954) has been used to explain the influence which environment may have to the formation of individual's self-beliefs. Social comparison processes offer a frame of reference (see Chapter 2.2.1), which enables individuals to reflect themselves in relation to others and therefore these comparisons are important building blocks for the individual's self-perceptions (see the triangle in Figure 3).

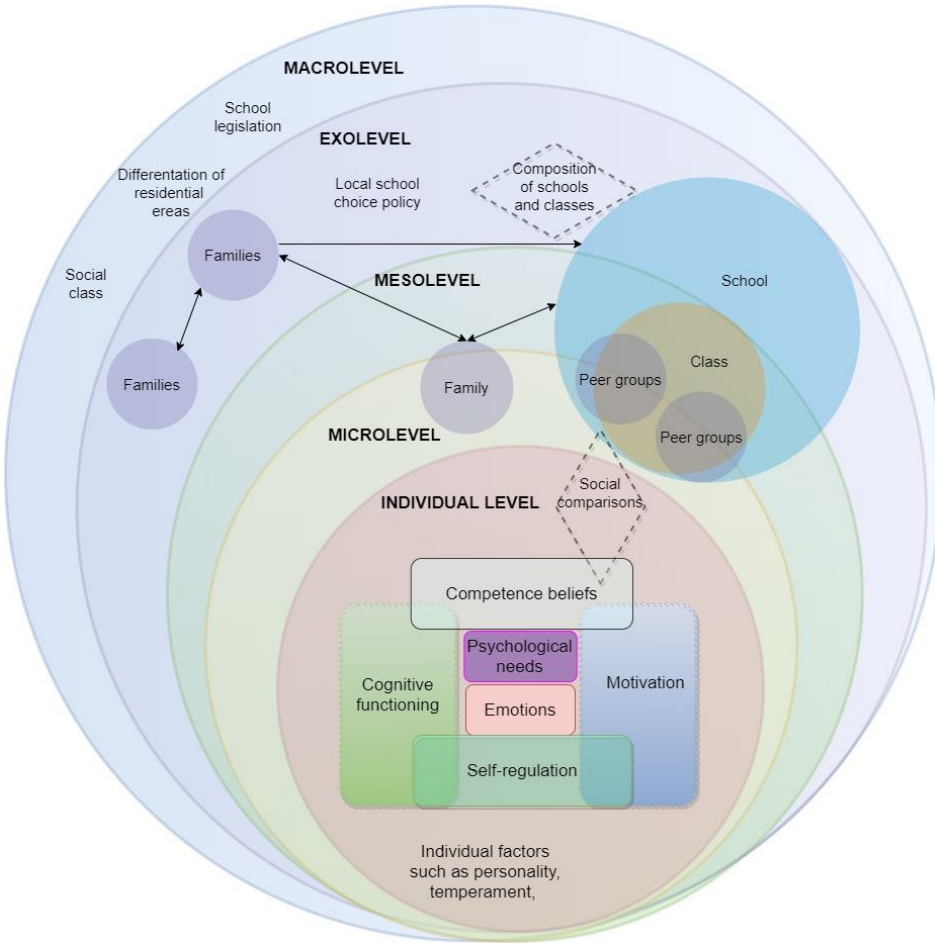

Figure 3 Summary of the key concepts and contexts of the research design, inspired by Bronfenbrenner (1977, 2005)

3 AIMS AND METHODS

3.1 Main aims

This dissertation is based on four original publications, which have been published (or accepted for publication) in peer-reviewed academic journals. These original sub-studies are referred to in the text by their Roman numerals (Studies I–IV).

The overall aim of this research was to investigate whether studying in a selective class with a special emphasis influences the development of pupils' competence related self-beliefs (i.e., action-control beliefs in studies I and II or academic self-concept in studies III and IV) or mathematical thinking skills (study III). In order to reach this overall aim, each sub-study examined it from a different perspective, which can be summarized into four research questions:

1. How do pupils' competence beliefs (studies I-IV) and mathematical thinking skills (study III) develop during the comprehensive school years?

2. Do classes with a special emphasis differ from classes without a special emphasis in terms of background factors (pupil's prior achievement, gender and mother's educational level) (studies I-III)?

3. Are there differences in the level of and change in pupils' competence beliefs or mathematical thinking skills between classes with and without a special emphasis (studies I-III [10]) after the initial differences due to selection process have been taken into account[11]?

4. Are peer effects on academic self-concept (Big-Fish-Little-Pond effect and Reflected Glory effect) or achievement (Peer Spillover effect) visible in the Finnish context and how is class type related to them (studies III and IV)?

[10] In studies I and III, the aim was to compare classes with and without a special emphasis, where classes with a special emphasis were a combination of all types of special classes. In study II, the comparisons were made between five different types of class: 1) regular class 2) arts (music or visual arts) 3) language 4) science and 5) sports.

[11] In each of the sub studies I-III, the focus was to study the value-added effect of class type after the confounding effects of independent variables (mother's education level, gender and prior achievement) were taken into account. In other words, the focus was to examine whether the possible differences between class types were explained by the selection process to the classes with a special emphasis (see Chapter 2.5). However, in sub study III the differences in background factors in different types of class were also examined in more detail.

3.2 Data

The data used in this study were drawn from a longitudinal learning-to-learn study collected by the Centre for Educational Assessment at the University of Helsinki, where children's cognitive abilities and motivational beliefs were assessed at several measurement points throughout the comprehensive school years in one large municipality in Finland.[12]

3.2.1 Data collection procedure and ethics

Data collection[13] started in autumn 2007 in a large city in southern Finland. The goal was to get a representative sample of the first-graders in that city. To ensure the representativeness of the sample schools, the selection was done using the equal-probability method. The aim was to get a sample of 800 pupils and sampling proceeded until there were enough pupils and schools in the sample. Therefore, at the beginning, 17 schools were included in the sample. However, out of those, one refused to participate and therefore the final number of schools at the beginning of the study was 16. The sample size was increased later at the beginning of fourth, seventh and ninth grade due to children's school choices and changes (see Figure 4 for the flowchart of the sampling procedure). As the purpose of this study was to examine the effects of class type on pupils' learning, the general principle in the data collection was to follow intact classes. Therefore, when pupils changed classes, new classes were added to the sample when at least four pupils from the original follow up went into the same class.[14] Therefore, at the beginning of fourth grade, four new schools with six classes were added to the sample. At the beginning of seventh grade, the sample size was increased again as most pupils changed schools.[15] In the seventh grade, 150 classes from 29 schools participated in the study. In ninth grade, all pupils in the city being examined were included in the sample (i.e., 50 schools). The numbers of pupils in each measurement point are presented in Table 1.

At each measurement point, the main responsibility for the data collection was given to classroom teachers who collected the data during regular school hours. At the beginning of the study, project leaders from the research team visited some

[12] The data collection was not originally planned for the purposes of this thesis. However, I had the privilege to start my PhD in this project at the beginning of 2013 i.e., before the data collection at sixth grade. Therefore, the aims of my thesis were considered in the questionnaires and in the sampling procedure from sixth grade onward.

[13] The data collection procedure has been described in detail in a doctoral thesis by Mari-Pauliina Vainikainen (Vainikainen 2014). See also (Lönnqvist et al., 2011; Vainikainen et al., 2015).

[14] Individual pupils who changed classes were followed up only separately.

[15] At the transfer stage from primary school to lower secondary school, most of the pupils had to change schools as only some of the schools are joint comprehensive schools including grades from one to nine within the same school. For example, in 2019 approximately 20% of all comprehensive schools in Finland were joint comprehensive schools (OSF, 2020).

of the schools for quality monitoring purposes. To minimize the missing information during the followed years, the research coordinator of the project had close contact with the schools and followed up on children over the years. She was responsible for following up on children, including those who had changed schools, and she visited the schools on multiple occasions and collected some of the data herself. Therefore, the number of missing participants between measurement points has been relatively small (see Appendix A for detailed information).

During the research process, ethical standards described by the Finnish Advisory Board on Research Integrity (see, TENK, 2012) were met. At the beginning of the study, municipal school authorities reviewed the research proposal and granted permission to collect the data. Parents were informed about the study through the Education Department of the city and families of all children were sent an information packet providing a summary of the project as well an agreement to participate in the study. Participation was voluntary and all families had a chance to refuse to participate at any point in the follow up. Anonymity of responses has been maintained at all stages of the study, as researchers have had access to de-identified data. Only data managers have had access to a separate database in which personal details have been stored.

3.2.2 Participants

For the purpose of this thesis, data mainly from classes from fourth to ninth grade were analyzed. However, information from grades one and five were also used in the study IV (e.g., school readiness test score from grade one and sociometric nominations from grade five, see Chapter 4.4. for more information). Descriptive information of the participants is presented in Table 1.

Table 1 *Descriptive information of the participants*

Grade	n	Gender		Mean age	Class type	
		girls/boys%		years	Class without an emphasis	Class with a special emphasis
1st grade	744	52 %	48 %	6.95	-	-
4th grade	1025	52 %	48 %	9.60	29 %	71 %
6th grade	1058	52 %	48 %	12.27	-	-
7th grade	2339	51 %	49 %	12.97	52 %	37 %
9th grade	4079	51 %	49 %	15.25	-	-

NOTE: Class type was collected only at grade four in primary school and grade seven in lower secondary school. In lower secondary school, in addition to the amount of pupils listed in the table, approximately 7% of the pupils (n=182) studied in mixed classes where special emphasis teaching was provided within a normal class and approximately 2% (n=41) of pupils studied in special educational needs (SEN) classes. These classes were not included in the analyses.

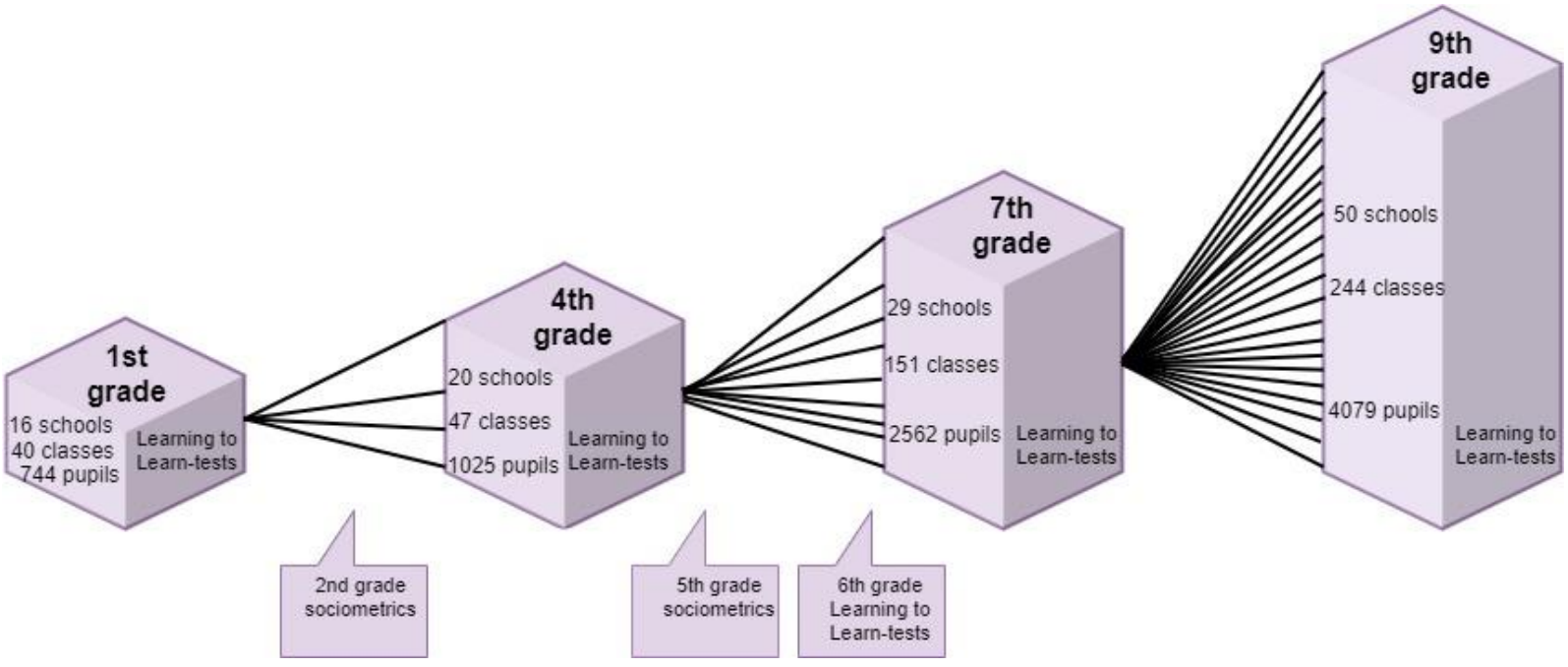

Figure 4 Flowchart of the sampling

Table 2 *Summary of the measures in different studies*

	Learning-to-Learn 1st grade	Parents' questionnaire 4th grade	Teacher's questionnaire 4th grade	Learning-to-Learn 4th and 6th grade	Learning-to-Learn 7th and 9th grade	Additional
Study I		mother's education	class type evaluation of achievement in 4 school subjects	action-control beliefs		
Study II					action-control beliefs mother's education achievement test score (i.e. Missing-Premises Tasks; Hidden Arithmetical Operations Task)	class type information from directors of schools in grade 7
Study III		mother's education	class type	mathematical self-concept mathematical thinking skills test (i.e., Mental Arithmetic Test; Hidden Arithmetical Operations Task)		
Study IV	school readiness test	parent's education	evaluation of mathematics achievement	mathematical self-concept mathematical thinking skills test (i.e., Mental Arithmetic Test; Hidden Arithmetical Operations Task)		sociometrics from grade 5

3.3 Measures

The data consisted of several measurement points and tasks and were derived from a longitudinal learning-to-learn assessment. Finnish learning-to-learn assessments include both cognitive and affective evaluations of children's learning-to-learn capabilities (Hautamäki et al., 2002). Cognitive tests include measures of mathematical thinking skills, reading skills and more general thinking and reasoning skills. Cognitive competences are assessed with tasks that are related to, but not directly linked to, curricular contents (Hautamäki et al., 2010; Hautamäki & Kupiainen, 2014). Affective factors include self-evaluation scales on several motivational concepts including action-control beliefs and academic self-concepts. In addition, during the assessments, complementary information was collected from parents and teachers.

Table 2 summarizes measurement points and measures that are relevant for the purpose of this thesis. The correlation matrices of all variables in different sub-studies are presented in Appendix B. Means and distributions of all items are presented in Appendix C.

3.3.1 Dependent variables

Competence beliefs

Competence beliefs were measured according to two theoretical frameworks. In studies I and II, the theoretical framework was action-control theory (Skinner et al., 1988, 1990, 1998) and in studies III and IV, competence beliefs were defined in the framework of academic self-concept by Marsh and his colleagues (e.g., Marsh, 1984; Marsh et al., 1988; Marsh & Craven, 2006; Marsh & Martin, 2011). The items of each scale are presented in Appendix C.

In study I, only *agency beliefs of abilities* and *agency beliefs of effort* and *control expectancy of learning* were the focus, whereas in study II all three belief types from action-control theory were included. Therefore, measures in study II included belief scales regarding *agency beliefs of ability* and *agency beliefs of effort, control expectancy of learning* and *control expectancy success* and *means-ends beliefs of luck* and *means-ends beliefs of abilities*. Each belief scale was assessed with three items using a seven-point Likert scale ranging from one (not true at all) to seven (very true).

Mathematical self-concept was assessed in studies III and IV with a scale based on Marsh's work on academic self-concept (see Marsh et al., 1988; Marsh et al., 2017). The scale consisted of three items on a seven-point Likert scale ranging from one (not true at all) to seven (very true).

Mathematical thinking skills

In study III, the change in children's *mathematical thinking skills*[16] was explored with measures that consisted of two tests. In the Mental Arithmetic test (modified from the Arithmetic subscale of the Wechsler Adult Intelligence Scale - Revised WAIS-R: Wechsler, 1981), the teacher read aloud a verbal mathematical problem and pupils answered in test booklets. Problems were for example: "If you buy two bus tickets and one ticket costs 3 euros 50 cents, how much money do you get back if you give 10 euros?" In Hidden Arithmetical Operators (Demetriou et al., 1991, 1996), pupils had to find hidden operators to solve the equation. In each item, there were one to four hidden operators (e.g., [(5 a 3) b 4 = 6. In this task letter a / b stands for: addition (+) / subtraction (-) / multiplication (•) / division (÷)?]). Each item was coded dichotomously as wrong (0) or correct (1) and the mathematical thinking skills test score was an average score based on five items from the Mental Arithmetic test and four items from the Hidden Arithmetical Operators test. Identical items were presented to pupils at both measurement points (grade 4 and 6). The total score had reasonable reliability at both measurement points (for grade 4 $\alpha = .65$ and for grade six $\alpha = .68$).

3.3.2 Independent variables

In all sub-studies, the analyses were performed on independent variables that were added to the models in order to take into account the original differences between pupils in different types of class. The independent variables used in the analyses were gender (all studies), class type (studies I-III), family background (i.e., mother's education level in studies I-III or parental education level in study IV) and prior achievement (all studies).

Gender, family background and class type

Gender (boy or girl) information was collected from the pupils and missing information at some measurement points was added from other available points.

Information on *class type* was collected from school administration personnel at the primary school. For lower secondary school data, class type was sought on pupils' assessment forms and missing or ambiguous answers were complemented and confirmed by the directors of each school. School directors were also asked some complementary questions regarding allocation of classes and based on this information classes where emphasized teaching was provided within regular classes (mixed classes) were excluded from the analyses (see Chapter 3.2.2). In studies I and III, class type was examined by comparing classes with and without

[16] For a more detailed description of the cognitive tests used in this specific learning-to-learn assessment, see Vainikainen, 2014.

a special emphasis, whereas in study II the options were 1) class without a special emphasis 2) arts (music or visual arts) 3) language (with different language options) 4) science and 5) sports).

Mother's education level was collected at the primary school level from parents (for the purpose of this study, at the fourth-grade measurement point) and at the lower secondary school level from the pupils themselves. At the primary school level (studies I and III), mother's education was collected using a scale with four options: 1) comprehensive school 2) general or vocational upper secondary school, 3) post-secondary education or higher vocational level, polytechnic education or bachelor's degree, and 4) master's degree or higher. At the lower secondary school level (study II), mother's education level was obtained using a scale with five options: 1) comprehensive school, 2) vocational upper secondary school, 3) general upper secondary school, 4) postsecondary education or higher vocational level, polytechnic education or bachelor's degree, and 5) master's degree or higher.

Parental education level was used in study IV to control for differences in pupils' family background. The variable was a combination of mother's and father's education level collected from the fourth-grade assessment from parents. The variable was computed so that it considered the highest level of either mother's or father's education.

Prior achievement

P*rior achievement* was used as an explanatory variable in all studies, but the variable that was used varied in each study.

In study I, *prior school achievement* was a composite score (GPA) based on the teacher's evaluation of the pupil's achievement in mother tongue and literacy, mathematics, science and foreign language. These scores were reported by teachers at the beginning of grade four (M=8.20, SD=.95) on a scale ranging from 4 [failed] to 10 [excellent]).

In study II, *prior achievement score* was derived from the seventh-grade assessment. The test score was a combination of two tests. The first consisted of eight items from the Missing Premises task of the Ross Test of Higher Cognitive Processes (Ross & Ross, 1979). Pupils were asked to make the given conclusion (e.g., Conclusion: Lake Saimaa is too cold for swimming) valid by giving them first one fact (e.g., First fact: The temperature of Lake Saimaa is 5°C). After that, they had to choose a second fact from five options (e.g., Most lakes are too cold for swimming.; It is winter.; Water that is 5°C is too cold for swimming.; Lake Saimaa is always cold.; Swimming in cold water is no fun.) that would make the given conclusion valid. The second test was the same Hidden Arithmetical Operators task (Demetriou et al., 1991) as in study III (see Chapter 3.3.1), but in study II, the test included items for older pupils and consisted of six items. All

items for each test were scored as correct or incorrect and the prior achievement test score that was used was a compositional score of all correct answers. The reliability of the test was reasonable ($\alpha = .65$).

In studies III and IV, the *mathematical achievement test score* used as an explanatory variable was based on the same mathematical thinking skills test score that was used as a dependent variable in study III (see Chapter 3.3.1). In both studies III and IV, this test score was used at different levels in the multilevel models (in study III, individual pupil's test score at the individual level and aggregated class averaged test score at the class level; in study IV, individual pupil's test score, aggregated peer group's average test score and aggregated class averaged test score).

In addition, in the final models in study IV, *teacher's evaluations of pupil's mathematics achievement in grade four* and *school readiness test score from grade one* were used in the analyses. *Teacher's evaluations of pupil's mathematics achievement in grade four* was the same as in study I, except that in study IV, only evaluation of mathematics achievement was considered. *School readiness tests from grade one* was a combination score from the first-grade assessment. It consisted of 17 items from three tests evaluating learning preparedness (for a detailed description of the test, see Hautamäki et al., 2001; Vainikainen, 2014) that measured children's capabilities to follow instructions[17], visual-spatial memory[18] and geometric reasoning skills[19]. Each of the 17 items was scored dichotomously as correct or incorrect and the test score was a compositional score of all correct answers.

[17] The pupils' capability to *follow the teacher's instructions* was assessed by a task measuring both children's inductive reasoning and executive functions. The test was originally developed by Elkonin (Raigodorosky, 2008) (see Raigorodsky [Ed.)] 2008) and modified for the Finnish learning-to-learn framework by Hautamäki and colleagues (2001). In this task, the pupils had to draw a path on an empty 12x5 grid according to the teacher's verbal instructions.

[18] *Visuo-spatial memory* was assessed by a task originally developed by Wilson, Scott and Power (1987) and modified by Logie and Pearson (1997). In the task, the pupils were shown different types of figures in a grid for three seconds and after seeing the picture, pupils were asked to reproduce the figure they saw on an empty grid.

[19] The Geometric analogies task was adapted from a Dutch geometric analogies test (Hosenfeld et al., 1997). In this task, the pupils were presented with a pair of geometric figures, e.g., a small square on the left and a big square on the right. The task was to apply the same rule when the pupil had to choose a pair from five options for another figure (e.g., a small circle). The transformations included adding an element, changing sizes and positions, halving and doubling, and the maximum number of simultaneous transformations was three.

Sociometric variables

In study IV, information regarding children's peer groups was drawn from sociometric tasks which children completed in fifth grade. In these tasks, pupils were asked to nominate up to five other pupils from within their class with whom they 1) worked with on school tasks, 2) played with during school breaks and 3) played with after school. These nominations were used to establish the peer group level for the multilevel analysis in study IV (for a detailed description of social network analyses, see study IV).

3.4 Data analysis

Data were analyzed with two statistical programs: SPSS (version 21.0 for study I, version 24.0 for studies II and III; version 25.0 for study IV) and Mplus (version 7.1 for studies I and II and version 8.0 for study III). In all sub-studies, descriptive information and missing values were first examined in SPSS. After that, measurement invariance was checked in Mplus.[20] Next, specific data analyses were used in order to answer the research question in each sub-study. In all the sub-studies, the research design was pretty similar, even though each study exploited different analysis methods. Figure 5 illustrates the general analysis model.

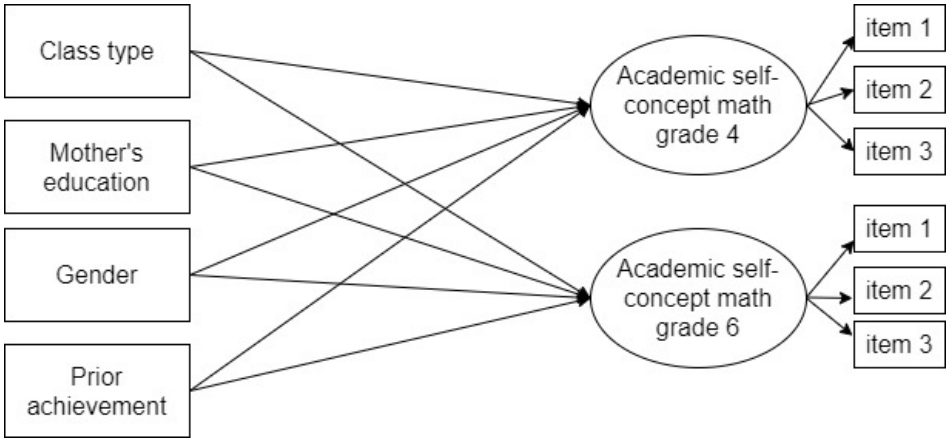

Figure 5 Illustrative example of the analysis models in different studies

Each of the first three sub-studies of this thesis examined the differences between class types. The focus was to examine both the level differences of the

[20] Measurement invariance was tested in studies I-III, in which one of the purposes was to examine the change of the factor means over time.

dependent variables (i.e., competence beliefs or mathematical thinking skills) in different types of class as well as differences in the development. Therefore, all three studies included either competence beliefs or mathematical thinking skills as dependent variables at two measurement points (either as latent factors or sum scores). In all models, the effects of class type were considered after the effects of the independent variables (i.e., gender, mother's education level and pupil's prior achievement) were taken into account. These specific covariates were added to models, as they have been noted to be related to the selection process of classes with a special emphasis (see Chapter 2.5).

Additionally, in the multilevel models of study III, the design resembled the design of the first three sub-studies, except that analyses proceeded at two levels (individual and class) with aggregated independent variables. In study IV, analyses were carried out with two or three levels (individual, peer group and class). Class type was not included in the analyses, but final models included more covariates.

3.4.1 Analyzing longitudinal data: measurement invariance and missing values

In order to make accurate interpretations from longitudinal data, it is important to ensure that the constructs being measured are comparable across different measurement points (Little, 2013; Widaman et al., 2010). It is possible that young children understand items in a survey differently than older children or adults. Therefore, before it is possible to make any interpretations about the change in some constructs or phenomena, it must be checked first that the structural validity of the measures is similar over the years. Therefore, prior to performing the actual analyses, measurement invariance over time was examined with longitudinal confirmatory factor analyses in Mplus. This made it possible to study whether the structure of the measures had been similar enough at different measurement points and whether the measures were valid for the actual analyses (Millsapp & Cham, 2013; van de Schoot et al., 2012). Testing of measurement invariance was conducted by hierarchically imposing restrictions on the model parameters and then comparing fit indices of different models in order to determine whether the model deteriorated or not (Widaman et al., 2010). Analysis proceeded from an unrestricted baseline model by first constraining the factor loadings (i.e., metric invariance) and then the intercepts of the items (i.e., scalar invariance) (Millsapp & Cham, 2013). In study I, examinations of measurement invariance were expanded to test the equality of residual variances (i.e. full uniqueness) (van de Schoot et al., 2012), but in studies II and III investigations were stopped after scalar invariance, which have been seen as a prerequisite for comparing factor means at different time points (Millsapp & Cham, 2013; van de Schoot et al., 2012).

Comparisons between different models were made by comparing fit indices of different models. The cut values that were used were: Comparative Fit Index (CFI) >.95, the root mean square error of approximation (RMSEA) <.06, and the standardized root mean square residual (SRMR) <.08 (Kline, 2005). In addition, for assessing comparative model fit, chi-square difference tests were performed (Millsapp & Cham, 2013). As the chi-square difference test is sensitive to sample size, changes in CFI of different models were also examined. According to Cheung and Rensvold (2002), a value of ΔCFI smaller than or equal to -.01 indicates that the assumption of invariance should not be rejected. Analysis of invariance showed that a sufficient level of invariance was achieved in each of the studies (see the original publications for detailed information), which made it possible to continue the analyses.

In addition to measurement invariance, analyzing and handling missing data are important in longitudinal data, where missingness occurs both between measurement points at the participants' level but also at the item level. In this study, due to the nature of the data collection from intact classes (see Chapter 3.2) and thanks to the active participation of the research coordinator, the number of missing participants between measurement points was relatively low (see Appendix A). At the item level, the randomness and the percentage of the missing data was examined with Little's MCAR test (Little, 1988). After this, adequate techniques for handling the missing data were assigned. In studies I and III, missing information was imputed with full-information maximum likelihood (FIML) estimation as Little's test showed missingness to be completely random (Cham et al., 2017; Wang & Wang, 2012). In study IV, a multiple imputation based on all variables that were used in the analysis was executed and the pooled estimates (see, Rubin, 1987) of all 50 imputations were used in the analyses. In study II, missing information was handled with listwise deletion as Little's MCAR test was significant.

3.4.2 Structural equation models (studies I and III)

In studies I and III, the analyses were carried out with structural equation models (SEM) in Mplus. Structural equation modeling (SEM) is a general term for a combination of statistical techniques (Kline, 2005; Lei & Wu, 2007; for a brief review of history of SEM, see Bollen, 1989/2014). Common for analysis in the SEM framework is that they enable analysis of both the measurement model and the path model at the same time (Lei & Wu, 2007; Silva et al., 2019). Therefore, SEM builds on factor analysis and multiple regression analysis and makes it possible to examine both the measurement model of latent constructs with many observed indicators as well as the relations between multiple predictors and outcomes at the same time (Silva et al., 2019). One of the main advantages of the SEM approach is that it can lead to more accurate estimations of the

phenomena being studied, as it reduces the measurement error in two ways. First, it takes into account the error in the observed variables and second, it enables fitting complex models with many predictors and outcomes (either observed or latent) only once, instead of through many separate analyses (Kline, 2005; Wang & Wang, 2012). Another benefit of the SEM approach is its flexibility (Lei, &Wu, 2007; Kline, 2005). SEM allows many types of analyses with fewer restrictive assumptions then traditional statistical analyses (Bollen, 1989/2014; Little, 2013). SEM applications also allow researchers to test, specify and adjust many of the assumptions (Little, 2013). Therefore, the general assumptions of SEM concern only sample size and normality. The SEM approach requires a large sample size (Kline, 2005; Lei & Wu, 2007) and many of the estimation methods assume that the distribution of variables is multivariate normal (Little, 2013; Kline, 2005). [21] In this thesis, the SEM approach has been used as the main analysis method in studies I and III. [22]

In study I, the analyses comprised first defining and examining the measurement model of the latent factors at two time points with longitudinal confirmatory factor analyses and examining the measurement invariance between factors at different time points (i.e. grade four and six) (see Chapter 3.4.1). After these structural validity checks of the measurement models, the change in latent means was explored to examine the overall change of action-control beliefs. Then, separate models were estimated for each dependent variable (i.e., agency beliefs of ability, agency beliefs of effort, and control expectancy). Each model included one belief scale from grade four and six as well as independent variables (i.e., class type, mother's education, gender and prior GPA) added as predictors (for an example, see Figure 3). This made it possible to examine whether class type explained the level differences in each belief type at the first measurement point (grade four) or the change in them (effects on grade six after controlling grade four) when the confounding effects of all independent variables were taken into account.

In study III, the analytical approach was similar except that comparisons between groups were done by using the multiple group option in Mplus (Muthén & Muthén, 2018). This enabled an examination of whether the relations between variables were similar in different groups (i.e., different types of class) (Little, 2013; Wang & Wang, 2012; Silva, et al., 2019). Therefore, in addition to measurement of invariance testing over time, measurement invariance between groups was also examined. After that, the actual analyses were carried out in two stages. First, with only dependent variables (academic self-concept or thinking

[21] According to Little (2013) it is reasonable to expect variables to be multivariate normal if they are univariate normal. Therefore, before the actual analysis, the general descriptive information of all variables was checked (see Appendix C).

[22] In sub study II, measurement invariance of the latent factors was checked but after that, factor scores were saved and transferred to SPSS and analyses were produced there in GLM style of ANOVA (see Chapter 3.4.4).

skills) at two time points (i.e., grade seven and nine), followed by independent variables (prior achievement, gender and mother's education). This made it possible to analyze whether the detected differences held after controlling for the effects of other independent variables. In order to increase the reliability of the estimates, analyses were carried out with the bootstrapping option, where confidence intervals for the estimates were bootstrapped with 1000 replicates (Wright et al., 2011). The hierarchical structure of the data (e.g., pupils nested within classes), was taken into account by the type complex option in Mplus (Muthén & Muthén, 2018).[23]

3.4.3 Multilevel models (studies III and IV)

In studies III and IV, analyses included more detailed examinations of compositional effects (i.e., Big-Fish-Little Pond, Peer Spillover and Reflected Glory effect). Therefore, multilevel modeling was exploited in these studies. In study III, analyses were carried out with latent factors by using multilevel structural equation models (MSEM) in the Mplus environment and in study IV, multilevel models with factor scores were performed in SPSS.

Multilevel models are advantageous methods for analyzing compositional effects, as they enable the study of effects of independent variables on the dependent one at several levels (i.e. school or class)(Finch & Bolin, 2017; Silva et al., 2019). In studies III and IV, the aim was to examine how class-level achievement predicted individual a pupil's achievement (study III) or academic self-concept (studies III and IV). Multilevel models can take into account the nested data structure and therefore correct for the non-independence of observations (Little, 2013). In other words, they consider that some pupils in the sample resemble each other more than others, as they come from the same schools or classes (i.e., in study III, the between level in the analysis was class and in study IV, the between levels were either peer group or class or both). In the analyses, all independent variables were standardized in order to make the interpretation of the results easier. Analyses proceeded with robust maximum likelihood estimator (MLR) (Muthén & Muthén, 2018) in studies III and with restricted maximum likelihood (REML) in study IV (SPSS Inc, 2005). As classes that were included in the analyses were mainly complete classes, manifest aggregation with grand mean centering was used (Marsh et al., 2009) and cross-level effects were calculated straight from the between-level effects. All analyses were first

[23] Intraclass correlation coefficients (ICC) varied depending on the variable (see Appendix D). As can be seen, variance between classes was relatively small for all other variables but in mathematical thinking skills, the variance between classes was larger, indicating that class membership explained part of the variance in mathematical thinking skills (Finch & Bolin, 2017). Therefore, the type complex-option in Mplus was used, as it takes into account the dependency in observations due to cluster membership (Muthén & Muthén, 2018).

executed without independent variables and after the first analyses, independent variables were added step by step to the models.

3.4.4 Mixed model analyses of variance (study II)

In study II, the analyses proceeded in SPSS with GLM (general linear model) style repeated measures analyses of variance (ANOVA). In the analyses, the dependent variable was one of the action-control beliefs factor scores[24] at two time points (within-subjects factor) and class type was a grouping variable (between-subjects factor). Therefore, analyses could also be described as a mixed model or split-plot type of ANOVA (Kraska, 2010; Murrar & Brauer, 2018), as it included both within- and between-subjects factors.

Repeated measures ANOVA takes into account the non-independency of error terms between different time points, but other general assumptions (i.e. normality and homogeneity of variance) should still be met (Field, 2017; Judd et al., 2017). For mixed models, these assumptions of normality and homogeneity of variance should be met both between and within groups (Kraska, 2010). As analyses included only two measurement points, the homogeneity of variances within different time points (i.e., sphericity) was not examined. Analyses of residuals showed them to be approximately normally distributed, but homogeneity tests revealed that variances between groups were not similar in all variables.[25] As unequal variances can violate p-values when group sizes are unequal, group sizes were equalized by taking a random sample of the biggest group (i.e. class without a special emphasis) (see, Field, 2017; Zimmerman, 2004). The goal was to get 20% of pupils from the original group in the new sample, so that group sizes would be roughly equal. However, random sampling proceeded at the class level in order to get comparable groups for the analyses.[26]

The analyses proceeded in two stages: first with only class type as a grouping variable, followed by background variables (gender and mother's education level as between-subjects factors and pupil's prior achievement as a covariate) that were included in the model so that their effects on the dependent variables would be taken into account. It was then possible to study whether the detected differences between classes were explained by background variables, and therefore were due

[24] Factor scores for each set of agency beliefs were saved and transferred to SPSS after measurement invariance testing (for a more detailed description, see study II).

[25] As Levene's test of homogeneity has been criticized when sample sizes are unequal (see Field, 2017), homogeneity of variances between groups was also examined with the non-parametric Levene's test (Nordstokke & Zumbo, 2010), which showed that variances between groups were unequal regarding some of the action-control beliefs.

[26] Other groups were already relatively similar in size and therefore random sampling only proceeded with the class without a special emphasis group. As all other classes were intact classes, it was important that classes in the "without a special emphasis" group would also be whole classes instead of random samples at the pupil level.

to the selection process of the classes with a special emphasis, or if they were actual "value-added" differences that occurred over the explanatory variables.

As the main aim was to find out both the level differences between class types in the dependent factors, but also the change in the dependent factors between different time points, the interaction term between time and class type was also examined. This made it possible to examine whether the change in dependent variables over time was different in different classes. Therefore, in the analyses, the focus was to examine three effects: 1) the main effect over time (i.e., the overall change in action-control beliefs), 2) the main effect for groups (i.e., differences between class types) and 3) the interaction effect between time and groups[27].

[27] In the analyses, information about the effects of background variables was also received, but they were excluded from the examinations, as they were only included in analyses in order to control the effects of class type.

4 OVERVIEW OF THE ORIGINAL STUDIES

The main goal of this dissertation was to analyze how pupils' competence-related self-beliefs and thinking skills develop during comprehensive school and to examine whether this development was similar between pupils studying in classes with and without a special emphasis. This goal was addressed in four empirical sub-studies. In this chapter, I will present the main findings of each of the original studies. Further details are available in the original publications. The main results from studies I-IV are summarized in Table 3.

4.1 Study I

Koivuhovi, S., Vainikainen, M.-P., Kalalahti, M. & Niemivirta, M. (2019): Changes in Children's Agency Beliefs and Control Expectancy in Classes With and Without a Special Emphasis in Finland from Grade Four to Grade Six, *Scandinavian Journal of Educational Research*, 63 (1), 427–442.

The first study drew on the action-control theory of Skinner and her colleagues (Skinner et al., 1988, 1990, 1998; for a more detailed description of action-control beliefs, see Chapter 2.1.1) and examined the change in children's agency beliefs and control expectancies in classes with and without a special emphasis during the late primary school years in Finland. The aim was to explore how children's agency beliefs of ability and control expectancy change from grade four to six and explore whether studying in a class with a special emphasis moderates this change after controlling for the effects of gender, prior school achievement and mother's education. In addition, the relations between explanatory variables were examined. It was expected that children who study in classes with a special emphasis would have a higher GPA (hypothesis 1 H1), more highly educated mothers (H2) and stronger self-beliefs (H3) than children in regular classes. In addition, it was expected that a higher prior GPA would predict positive self-beliefs (H4) and that gender would be associated with action-control beliefs differently depending on the type of belief. Girls were expected to display stronger ability beliefs of effort than boys (H5a), whereas boys were expected to have stronger agency beliefs of ability (H5b). Overall children's action-control beliefs were expected to decrease from grade four to six.

Participants (N=1025) came from 47 classes, of which 11 were classes with a special emphasis (n=291) (for a more detailed description of the data collection and participants, see Chapter 3.2). Six of the classes with a special emphasis were language classes, four focused on music and/or dance and one on sports. Thus, it

was not possible to consider different classes separately, but only to focus on the overall status of studying in a class with a special emphasis or in a regular class i.e., without an emphasis. The data were analyzed with structural equation models in Mplus. Analyses were done in two stages: first the measurement invariance of the latent factors were checked, then the actual analyses proceeded.

The main finding of study I was that class type did not predict either the change or the level of pupils' action-control beliefs when other background variables (gender, prior school achievement and mother's education) were taken into account. As expected, class type correlated with prior achievement, mother's education level and gender, i.e., studying in a class with a special emphasis was more typical for girls than for boys and for children with a high GPA and well-educated mothers.

Prior achievement (GPA) was the strongest predictor of self-beliefs and it predicted both agency beliefs of ability and effort and control expectancies in grade four. Moreover, prior achievement (GPA) predicted the change in agency beliefs of ability (but not effort) and control expectancies. Mother's education and gender were also significant predictors of action-control beliefs, but their effects varied depending on the type of the belief. Children with highly educated mothers displayed stronger control expectancy than others and their agency beliefs of ability changed more positively. Regarding agency beliefs of effort, mother's education level was not a statistically significant explanator. Gender was associated statistically significantly with agency beliefs, but not with control expectancy. Girls had stronger agency beliefs of effort than boys in grade four. However, in agency beliefs of ability, boys' beliefs remained stronger from grade four to six when compared to those of girls.

Overall, the predictions between grade four and six action-control beliefs were all significant, indicating some degree of stability over time. The relative change in action-control beliefs was non-significant for agency beliefs of ability, whereas the change was statistically significant for agency beliefs of effort and control expectancy. As expected, children's agency beliefs of effort decreased from grade four to six. However, at the same time, their control expectancies increased.

4.2 Study II

Koivuhovi, S., Vainikainen, M.-P. & Kalalahti, M. (*2020*): The effect of Studying in Selective Classes on the Change of Pupils' Action-Control Beliefs during Lower Secondary School in Finland. *Scandinavian Journal of Educational Research.* Advanced online publication.

Study II continued within the same theoretical framework as studies I and examined pupils' action-control beliefs during lower secondary school. The measures in study II comprised all aspects of action-control beliefs and thus also included means-ends beliefs, which is the third part of the triangle in action-control theory (for a more detailed description of the action-control theory, see Chapter 2.1.1). The increased number of participants in the sample during lower secondary school (detailed description of the data and sampling procedure see Chapter 3.2), made it possible to examine the differences between different emphasized subjects instead of comparing two categorical groups as in study I.

Thus, the aims of study II were to analyze how pupils' agency and means-ends beliefs as well as control expectancies differ and the change in different types of class during lower secondary school years. The focus was on both the overall level differences between different types of class (classes without a special emphasis, language, arts, science and sports classes) as well as the differences in the change of those beliefs from grade seven to nine.

Participants (N=1839) came from 30 schools and 150 classes. Most of the pupils, 58% (n = 1077) of the participants, studied in classes without a special emphasis and the rest in classes with a special emphasis (14% (n = 251) languages, 11% (n = 194) arts, 9% (n = 171) science and 9% (n = 146) in sports classes). Before the actual analyses, the measurement invariance of the latent factors was examined in Mplus and factor scores were saved for the actual analyses. Analyses were conducted in two phases, beginning with only class type as a classifying variable and factor scores for each set of action-control beliefs as dependent variables. After that, other independent variables (mother's education, gender, prior achievement) were added to the analyses. This made it possible to examine whether the selection process into classes with a special emphasis could explain the differences detected between classes. In order to get comparable groups for the analyses, groups were equalized by taking a random sample of 20% (n = 238) of the biggest group (i.e., pupils in classes without a special emphasis) (for more information about the data analyses, see Chapter 3.4.4).

Results from study II showed that different class types clearly differed from each other regarding background variables. Language classes seemed to be the most selective as pupils studying in them had the highest GPA and most highly educated mothers. In addition, science and arts classes seemed to attract high-achieving pupils, whereas sports classes were most similar to classes without a special emphasis, but still differed significantly from them regarding achievement

in grade seven. Gender differences in the allocation to different class types were most pronounced in science and arts classes: science classes attracted significantly more boys than expected, whereas girls favored arts classes.

One of the main findings in study II was that when background variables were taken into account, most of the differences between classes disappeared. This result concerned both the level differences[28] in action-control beliefs between different types of classes but also the change[29] of them. Therefore, results indicate that for the most part, the change in pupils' action control beliefs was similar in different types of class. The general tendency was that positive action-control beliefs declined, whereas negative beliefs remained quite stable from grade seven to nine.

However, pairwise comparisons revealed some minor differences between classes. Interestingly, children studying in science classes seemed to undergo more positive changes than children in other classes. In science classes, pupils' control expectancies of success remained more positive than in other classes and at the same time, their detrimental means-ends beliefs of abilities decreased more than in other classes. This finding slightly supported the hypothesis of the beneficial effects of selective peer groups. However, regarding language and arts classes, results also provided evidence for divergent peer effects. In language and arts classes, pupils' agency beliefs of abilities declined more than in other class types. This finding was interesting given the notion that pupils studying in language and arts classes were the ones with high levels of achievement. Therefore, this finding indicated that pupils studying in language and arts classes might suffer from the Big-Fish-Little-Pond phenomenon (BFLPE) (Marsh, 1987; for a review, see Marsh, Seaton, et al., 2008; Marsh & Seaton, 2015). Similarly, findings regarding pupils' control expectancy beliefs of learning confirmed this interpretation, as the change in them seemed to be slightly more negative in language classes than in other classes.

[28] Regarding level differences, all level differences, except for one (i.e., control expectancy of learning), between classes disappeared when background factors were considered. In control expectancy beliefs of learning, however, level differences remained significant in both grades. Pairwise comparisons showed that pupils studying in science and language classes had significantly stronger control expectancies of learning than pupils in classes without a special emphasis in grade seven. However, in grade nine, only science classes differed from classes without a special emphasis and the difference between language classes and classes without a special emphasis was no longer significant. This indicated that the decrease in children's control expectancies of learning was slightly greater in language classes than in other classes (even though the interaction effect between time and class was non-significant).

[29] When background variables were considered, only one of the interaction effects between time and class type (i.e., regarding control expectancy of success) was statistically significant.

4.3 Study III

Koivuhovi, S., Vainikainen, M.-P. & Kalalahti, M. (*2021*). Oppilaiden matemaattisten ajattelutaitojen ja matematiikkaminäkäsityksen kehitys painotetun opetuksen ja yleisopetuksen luokilla neljänneltä luokalta kuudennelle. [Development of pupils' mathematical thinking skills and mathematical self-concept in classes with and without a special emphasis from fourth to sixth grade] *Kasvatus,* 52 (1), 22–36.

As some of the findings in study II showed some interesting signs of possible Big-Fish-Little-Pond effect (BFLPE), (e.g., Marsh & Parker, 1984), study III focused on academic self-concept, which is the field in which BFLPE was originally invented (Marsh et al., 1988; Marsh & Martin, 2011; Shavelson et al., 1976; see also Chapter 2.1.2). In addition to studying the development of academic self-concept in different types of class, study III also focused on achievement and peer effects. Thus, it examined three of the more common peer effects in educational research: the Big-Fish-Little-Pond effect (i.e. negative effects of class-average achievement on an individual's academic self-concept) and the Reflected Glory effect (i.e. a boost effect that pupils in selective tracks might get from knowing that their track is highly ranked and better than others) on academic self-concept as well as Peer Spillover effect (i.e. a positive effect of class-average achievement on individuals' achievement) on achievement (for a more detailed description of academic self-concept and each peer effect, see Chapters 2.1.1; 2.4.1 and 2.4.2).

More precisely, study III first examined how pupils' academic self-concept in mathematics and mathematical thinking skills develop in classes with and without a special emphasis from grade four to six. Secondly, it explored the effect of gender and mother's education level on mathematics self-concept and thinking skills. In addition, thirdly, it examined the three above-mentioned peer effects and aimed to find out how class type relates to them.

Participants (N=939) came from 47 classes, of which 11 were classes with a special emphasis (for a more detailed description of the data collection and participants, see Chapter 3.2). Six of the classes with a special emphasis were language classes, four focused on music and/or dance and one on sports. As in study I, a small number of pupils in different emphases limited the opportunity to compare different classes separately and thus classes were combined, focusing on the overall status of studying in a class with a special emphasis or in a regular class i.e. without an emphasis. Of all participants 71% (n=664) studied in classes without a special emphasis and 29% (n=275) in classes with a special emphasis. Compared to studies I and II, the analyses of study III proceeded in a multilevel setting, which took into account the hierarchical structure of the data and thus provided more detailed analyses than in studies I and II. Again, the measurement invariance of the latent factors was examined first before other analyses.

The results in study III showed that, as in studies I and II, the trend in the change in pupils' self-beliefs was declining and overall pupils' academic self-concept in mathematics declined from grade four to six. The decline was steeper in classes without a special emphasis, but when differences in background variables (gender, prior achievement, mother's education level) were taken into account, pupils studying in classes with and without a special emphasis did not differ from each other in terms of the strength or change in mathematics self-concept.

When pupils' mathematical thinking skills were considered without considering the background variables, it showed that mathematical thinking skills developed in both class types and that pupils in classes with a special emphasis scored higher on the test of mathematical thinking skills in both fourth and sixth grade. However, adding background variables to the model removed most of the differences. Interestingly, there was still a marginally significant difference between class types in the fourth grade after the background variables had been included. However, in sixth grade the difference between class types had disappeared. In addition, regarding the change in mathematical thinking skills, adding the background variables revealed that the increase in pupils' mathematical thinking skills from grade four to six was statistically significant only in classes without a special emphasis. In other words, pupils' mathematical thinking skills in classes with a special emphasis seemed to develop slightly less than would have been expected in terms of differences in background variables. Even though the difference in development was very small, the results indicated at least that studying in a class with a special emphasis did not boost pupils' mathematical thinking skills.

As expected, boys had a stronger self-concept in mathematics than girls. However, interestingly, the gender difference was statistically significant only in classes without a special emphasis, whereas in classes with a special emphasis, the gender difference was significant only in grade four, but no longer in grade six. In other words, in classes without an emphasis, boys believed in their mathematical skills more than girls did and gender had a significant effect on the change in mathematics self-concept from grade four to six, indicating new gender differences favoring boys. On the contrary, in classes with a special emphasis, the gender difference evened during the last years of primary school.

Mother's education level significantly predicted pupils' test score on the test of mathematical thinking skills both in grade four and six. In other words, children with highly educated mothers had better thinking skills in mathematics and their skills also developed more than children with less educated mothers. However, mother's education level was not a significant predictor of pupils' self-concept in mathematics.

Regarding peer effects, analyses showed that BFLPE was visible in the data, whereas RGE and PSE were not. Therefore, findings from study III suggest that

studying in a class with a high average achievement level may have detrimental effects on the individual pupil's self-concept.

4.4 Study IV

Koivuhovi, S., Marsh, H.W., Dicke, T., Sahdra, B., Guo, J. Parker, P.D., & Vainikainen, M.-P. (*2020*): Academic Self-concept Formation and Peer-Group Contagion: Development of the Big-Fish-Little-Pond Effect in Primary-school Classrooms and Peer Groups. *Journal of Educational Psychology*. Advanced online publication.

The fourth study of this thesis elaborated on the examinations of Big-Fish-Little-Pond Effect (Marsh, 1987; for a review, see Marsh, Seaton, et al., 2008; Marsh & Seaton, 2015). As results from study III had confirmed the existence of BFLPE in the Finnish context, the aim of the final study was to further examine this interesting peer effect.

The theory of BFLPE (i.e. negative effect of class-average achievement on an individual's academic self-concept) was initially formulated at the school level (Marsh, 1987; Marsh & Parker, 1984), but later studies have shown it to be tenable, and even more pronounced, at the class level (Guo et al., 2018; Marsh et al., 2015; Marsh & Seaton, 2015; Marsh et al., 2014). This finding is in line with the core assumption in local dominance theory (Zell & Alicke, 2010), which argues that individuals' concept of themselves is most strongly influenced by the most proximal reference group. However, despite the enormous number of studies in the field of BFLPE, few prior studies have explored the phenomenon on a more local level than class level. Therefore, the main aim of study IV was to extend the analyses of BFLPE and explore it simultaneously at the level of class and at the level of smaller peer groups inside classes. In addition, the change in mathematical self-concept from grade four to six was examined in relation to peer- and class-level achievement averages (i.e., BFLPE) and other covariates (i.e., teacher's evaluation of achievement, gender, parental education and school readiness test in first grade).

More specifically, it was expected that BFLPE would be present at both the peer group and class levels when those are examined separately. In other words, it was expected that the mathematics test score would predict mathematics self-concept at the individual level positively, but negatively at the peer group level (hypothesis 1) and class level (hypothesis 2) when peer group and class were examined separately. In addition, both peer group and class average achievement levels (when measured separately) were hypothesized to be negatively related to the change in mathematics self-concept. When both peer group and class levels were considered in the same three-level model, BFLPE was expected to be more pronounced at the peer group rather than the class level (hypothesis 3), as expected

in local dominance theory. In other words, class-average achievement was largely expected to be absorbed into the effect of peer group average achievement when both the peer group and class level were analyzed in the same model. Finally, it was expected (per hypothesis 4) that the detected result patterns would remain quite stable even after the effects of potentially confounding covariates (i.e., the teacher's evaluation of achievement, gender, parental education and school readiness test in first grade) were taken into account.

The data used in the fourth study were the same primary school data as those used in studies I and III (see Chapter 3.2). Therefore, the data consisted of 1017 pupils nested within 130 peer groups within 46 classes. Data were analyzed with several multilevel models in SPSS (version 25). Before the actual analyses, missing data were examined and accommodated with multiple imputation where 50 imputed data sets were constructed. Data analyses were then executed for the imputed data set and the pooled results were reported (Rubin, 1987). Multilevel models were carried out with following order: first, separate two-level models for peer group and class were carried out without any other predictors, then models were combined to form a three-level model to test the BFLPE simultaneously at both levels. After that, covariates (i.e., teacher evaluations of pupil's mathematics achievement, gender, parental education level and school readiness test score from grade one) were also added to the model. To study the change in mathematical self-concept, mathematics self-concept from grade four was also included in the model.

The results from study IV showed that, as expected (hypothesis 1 and 2), BFLPE was visible at both the peer group and class levels when they were examined separately. Interestingly, simultaneous analysis at the peer group and class level showed that BFLPE was more pronounced at the class than at the peer group level, which was contradictory to local dominance theory and our expectations (hypothesis 3). In the three-level model with both peer group and class level, the negative effect of the peer group average became non-significant, while the negative effect of class average remained relatively unaffected. As expected, (hypothesis 4), the patterns of the results remained stable even after the effects of covariates were taken into account. However, adding the covariates to the model revealed some interesting findings, especially regarding the change in mathematics self-concept from grade four to six. From the included covariates, teacher evaluations and gender significantly explained the change in mathematics self-concept from grade four to six, whereas the effects of parental education and the school readiness test were non-significant. In other words, teacher evaluations of a pupil's achievement in grade four influenced the pupil's mathematics self-concept in grade six, even more so than in grade four. Similarly, gender was associated with both mathematics self-concept in grade four and six separately, but also with the change in it. Boys had a significantly stronger mathematics self-concept in both years and the difference between genders intensified during the

two followed years. Overall, the analyses showed that class-level average achievement was related to the change in mathematics self-concept from grade four to six, which supported the growing role of social comparison processes during the primary school years.

Table 3 *Summary of the main findings of the original studies*[30]

Study	Main aims	Participants	Grades	Measures	Analyses	Main findings
1.	Explore the level of and change in agency and control expectancy beliefs in classes with and without a special emphasis	N=1025, 28% in classes with a special emphasis	4 to 6	Agency beliefs: effort and ability Control expectancy: learning	SEM in Mplus	Studying in a class with a special emphasis was more typical for girls than for boys and for children with a high GPA and well-educated mothers. Class type did not predict either the change or the level of pupils' action-control beliefs when background variables were taken into account.
2.	Explore the level of and change in action-control beliefs in different types of class with a special emphasis	N=1839, 14% in languages, 11% arts, 9% science, and 9 % in sports classes	7 to 9	Agency beliefs: effort and ability Control expectancy: learning and success Means-ends beliefs: luck and abilities	Mixed model ANOVA	Classes differed from each other in terms of background factors. Language classes seemed to be the most selective ones regarding GPA and mother's education. Science classes attracted significantly more boys than expected, whereas girls favored arts classes. The change in pupils' action-control beliefs was relatively similar in all class types. However, some detailed differences were found: children studying in science classes seemed to undergo more positive changes than children in other classes. In language and arts classes, certain beliefs developed more negatively than in other classes.
3.	Explore 1) change in mathematical self-concept and thinking skills in classes with and without a special emphasis and, 2) peer effects BFLPE, RGE and PSE in Finland	N=939, 29% in classes with a special emphasis	4 to 6	Mathematical self-concept and thinking skills	Multiple group and multilevel models in Mplus	The change in both mathematical self-concept and thinking skills were relatively similar in classes with and without a special emphasis. Studying in a class with a special emphasis seemed to protect girls' mathematical self-concept from a decline. BFLPE was explicit whereas RGE and PSE were not
4.	Explore BFLPE simultaneously at the peer group and class level	N=1017 pupils nested within 130 peer groups within 46 classes	4 to 6	Mathematical self-concept and thinking skills	Multilevel models in SPSS	BFLPE was visible at both the peer and class levels when they were analyzed separately. In simultaneous analyses, only the class level effect was significant. Boys had stronger mathematical self-concept than girls and the gender difference intensified from grade four to six. BFLPE intensified from grade four to six indicating the increasing role of social comparison processes during primary school.

[30] Acronyms which are used in the table are: GPA (Grade-Point-Average), BFLPE (Big-Fish-Little-Pond effect), RGE (Reflected-Glory effect, PSE (Peer Spillover Effect)

5 DISCUSSION

Discussion around classes with a special emphasis has been heated ever since their large-scale establishment in the 1990s (e.g., Rinne et al., 2017; Seppänen, Kalalahti, et al., 2015). This discussion is grounded on many unstudied assumptions and the focus of this thesis was to examine one of them. Classes with a special emphasis have become a popular policy in the urban areas of Finland, where especially well-achieving children from affluent family backgrounds choose to apply for them (e.g., Kalalahti et al., 2015; Kosunen, 2014; Kosunen & Seppänen, 2015). Therefore, classes with a special emphasis have been described as an "implicit tracking system" inside the Finnish comprehensive school (Berisha & Seppänen, 2017) and researchers have been worried about their effects on equality (Kosunen, Bernelius, et al., 2016; Varjo et al., 2014). However, at the same time, allowing pupils a chance to follow their individual interests within comprehensive school by choosing a class with a special emphasis has been considered important by local educational authorities, policy makers and families (see e.g., Seppänen, 2003; 2006; Varjo & Kalalahti, 2015). Especially parents have been eager to have the opportunity to choose a selective class, as they seem to believe that studying in a class with a special emphasis might be of benefit to their child and increase his/her motivation to do school work (Kosunen & Carrasco, 2016; Kosunen & Seppänen, 2015). Even though classes with a special emphasis and school choices have been studied for over a decade in Finland, no studies prior to this one have analyzed the effects they may have on pupils' motivation.

The main goal of this dissertation was thus to find out whether empirical evidence can support the assumptions regarding the beneficial effects of classes with a special emphasis. This goal was addressed through four empirical studies, each with different perspectives on the issue. Each of the studies I-III focused on analyzing the differences between classes with and without a special emphasis and examined how children's competence beliefs and mathematical thinking skills develop in different types of class. Then again, the final study (IV) elaborated the findings of the earlier studies and examined the mechanisms of peer influence in more detail. Each sub-study had its own specific research questions, which were summarized into four research questions for this thesis (see Chapter 3.1). Next, the main findings from this thesis will be discussed.

5.1 Main findings

5.1.1 Development of competence beliefs and mathematical thinking skills during comprehensive school

Before examining the differences between classes with and without a special emphasis, it was important to analyze the overall development of children's' competence beliefs and mathematical thinking skills during comprehensive school, so this was the focus of the first research question of this thesis.

Findings from all sub-studies were consistent with prior research (Eccles et al., 1993; Jacobs et al., 2002; Pintrich & Schunk, 1996; Vainikainen, 2014; Wigfield & Eccles, 2000; for a review, see Stipek & MacIver, 1989; Muenks et al., 2018) and showed a generally declining trend in the development of children's competence beliefs during primary and lower secondary school. These findings have been explained as being a result of developing cognitive processes, which lead to better self-awareness and more accurate self-evaluations (Demetriou et al., 2011). This maturing of cognitive processes was visible in study III, which showed that children's mathematical thinking skills progressed during the follow up.

In addition to general developmental and maturing processes, self-beliefs are affected by social comparisons, which shape individuals' concept of themselves. The importance of social comparison processes have been shown to increase alongside cognitive development (Dijkstra et al., 2008); on the other hand, some studies (Burke & Sass, 2013) have suggested that the role of social comparisons would be particularly strong in primary schools where children spend more time in the same classes than they do in lower secondary school. Based on the results from this thesis, it is possible to conclude that positive self-beliefs decline during the school years. The decline in competence beliefs was detected both in primary school (e.g., from grade four to six; studies I and III) as well as in lower secondary school (study II). Findings from study IV suggest that, indeed, the increasing role of social comparisons may be one explanation for this decline. Longitudinal analyses of BFLPE in study IV showed that the magnitude of BFLPE intensified during the two followed years. Prior cross-sectional studies (Marsh et al., 2015; Parker et al., 2019) that have examined the development of BFLPE have suggested that this increasing size of BFLPE could be partly explained by differences in the ability stratification at different school levels (Parker et al., 2019). In other words, studies have suggested that a more pronounced BFLPE in older children would be because older children are usually being taught in a more tracked school system. However, the longitudinal data set used in this study made it possible to examine the phenomenon in a context in which children stayed in the same classes

throughout the school years[31]. Findings suggest that at least in some part, the increasing magnitude of BFLPE would be attributed to developmental processes through which growing cognitive capabilities lead to greater importance of social comparisons.

Results from studies I and II confirmed the findings of prior research (Geldhof, Little, 2011; Little et al., 1999) and showed that even though the general trend in self-beliefs was declining, the trajectories of some specific beliefs were different. Both studies I and II showed a decline in children's agency beliefs of effort, but the decline in agency beliefs of abilities was explicit only in lower secondary school (study II), whereas the ability beliefs remained stable in primary school (study I). Previous longitudinal studies (Little et al., 1999) have reported similar findings. Results indicate that children's self-evaluations develop at different stages. Children's understanding of ability and effort start to differentiate only at the end of the primary school years (Nicholls, 1978), which explains the stability of ability beliefs detected in primary school (study I). Regarding control expectancies, the results showed a similar type of variation in the development between primary and lower secondary school. At the primary school level (study I), children's control expectancy for learning became slightly more positive during the follow-up period, whereas at the lower secondary school level, both control expectancies for learning and success decreased (study II). Findings may indicate that children's developing sense of autonomy and individuality in primary school is boosting their control expectancies and only after more experiences of achievement situations do their expectations become more realistic.

Means-ends beliefs and mathematical self-concept were each examined in separate studies (i.e., means-ends beliefs in study II and mathematical self-concept in study III) and therefore the period when they were followed was more limited than other beliefs. Regarding mathematical self-concept, the development was similar to other self-beliefs and children's overall mathematical self-concept decreased during the follow up period (study III). Then again, means-ends beliefs of luck and ability remained quite stable during the follow up period (study II), even though pairwise analyses showed some minor differences in the development between different class types (see Chapter 5.1.3).

The findings of this thesis also confirmed the reciprocal relationship between achievement and self-beliefs (e.g., Bong & Skaalvik, 2003; Guay et al., 2003; Marsh & Craven, 2006). In all sub-studies, achievement had a positive association with self-beliefs. The findings were thus in agreement with previous research showing that self-beliefs develop with experiences of success or failure. In other words, those who do better at school are also likely to believe their own competence more and vice versa (Bong & Skaalvik, 2003). In addition, results from studies I and IV showed that longitudinally the development seems to be

[31] In study IV, the period that was scrutinized was from grade four to six in primary school and therefore most of the children studied in the same classroom and peer environment during that time.

cumulative, as prior achievement also predicted the change in some of the self-beliefs (i.e., agency beliefs of ability and control expectancy of learning in study I and mathematics self-concept in study IV). Thus, the development of self-beliefs was more positive for well-achieving children.

Prior studies have shown that self-beliefs differ by gender (e.g., Butler & Hasenfratz, 2017; Malmberg et al., 2008; Marsh, 1989). Similarly, some typical gender differences were found in this study. Regarding action-control beliefs, gender differences were detected in agency beliefs (study I) and means-ends beliefs (study II), whereas gender was not associated with differences in control expectancies. Findings regarding agency beliefs were in line with previous studies (Malmberg et al., 2008) and showed that girls evaluated themselves as being more industrious than boys, whereas boys trusted in their own abilities more than girls (see study I).

Boys and girls also differed in their mathematical self-concepts (studies III and IV). As expected based on prior research (Marsh, 1989), overall boys had a stronger mathematical self-concept than girls. Gender difference also intensified during primary school (see study IV), but interestingly, results from study III showed that class type was confounded with gender, indicating that the gender difference in mathematical self-concept developed differently in classes with and without a special emphasis (see Chapter 5.1.3).

5.1.2 Selectivity of classes with a special emphasis

The second research task of this thesis was to examine how classes with a special emphasis differ from classes without a special emphasis in terms of background factors. As a reasonably large number of Finnish studies on school choice and classes with a special emphasis (e.g., Kosunen, 2014; Seppänen, Carrasco, et al., 2015; Seppänen, Kalalahti et al., 2015) have shown, the results of this thesis also showed that classes with a special emphasis are clearly selective by nature and appeal especially to well-achieving children from middle-class families. The selectivity of classes with a special emphasis was visible in all sub-studies in which it was examined (studies I-III). Children who studied in classes with a special emphasis had a higher GPA and their mothers were more educated than children in regular classes (i.e., class without a special emphasis). Similarly, international studies on school choices have shown that educational choices are especially important for middle-class parents who stress the importance of education and see it as an investment for the future (Ball et al., 2004; Gewirtz et al., 1995; Vincent, 2001; Vincent & Ball, 2007). Furthermore, middle-class families frequently manage to make more beneficial choices as they have wider social networks and recourses for choosing than less educated families (Ball & Vincent, 1998). In the Finnish context, the selection criteria of emphasized teaching has seen to favor middle-class children, as success in the aptitude tests

usually demands corresponding interests and prior hobbies, which, in turn, require more resources from the families (Kosunen & Seppänen, 2015).

Most Finnish studies on emphasized classes have compared classes with a special emphasis as such to classes without a special emphasis and only few studies have focused on the differences between different types of emphasis (but see Kupiainen & Hotulainen, 2019). In this dissertation, differences between different types of emphasized classes were analyzed in one sub-study (study II). The results revealed that classes with a special emphasis differed not only from regular classes, but also from each other in terms of background variables.

As prior studies have shown (Berisha & Seppänen, 2017), different emphasized subjects attracted boys and girls differently. In this study, gender differences were strongest in arts and science classes, where gender distribution clearly deviated from random. Girls were over-represented in arts classes and in turn, science classes attracted especially boys (see study II). A similar kind of gender difference has been reported before (Kupiainen & Hotulainen, 2019). However, interestingly, gender differences were less pronounced in other class types (i.e., language, sports) and non-significant, even though it could have been expected that, for example, sports classes would appeal especially to boys (Kupiainen & Hotulainen, 2019). The results therefore confirm the idea of "municipal school choice spaces" (Varjo & Kalalahti, 2011, 2019) noting that "who" chooses and "what" is regulated by the school choice policy of each municipality.

In terms of other background factors, all classes with a special emphasis differed from classes without a special emphasis in pupils' achievement. In other words, pupils' GPA was statistically significantly better in language, science, arts, and sports classes respectively, than in regular classes.[32] This finding was of course quite logical and expected and partly a consequence of the selection criteria for classes with a special emphasis (see Varjo et al., 2014; Varjo & Kalalahti, 2019). Additionally, classes differed from each other in terms of family background (i.e., mother's education). Pupils in language classes had better-educated mothers, but differences in family background were visible in all emphasized subjects when compared to classes without a special emphasis (see study II). Therefore, findings from studies I-III clearly confirmed the selective nature of classes with a special emphasis, which have been noted before in several studies (e.g., Kosunen, 2014; Seppänen, Carrasco, et al., 2015; Seppänen, Kalalahti et al., 2015).

[32] This result is based on data with equalized sample sizes. In the original sample, pupils in classes without a special emphasis had a higher achievement level (mean 5.08 compared to 4.68 in the new sample). Therefore, in the original sample, classes without a special emphasis did not differ from sports classes, whereas the difference was significant in the equalized sample. Regarding gender and mother's education level, both samples produced virtually similar results (for detailed information, see sub study II).

5.1.3 Differences between classes with and without a special emphasis in children's competence beliefs and mathematical thinking skills

The main task of this thesis was to examine whether studying in a selective class without a special emphasis influences children's motivation. Therefore, the primary focus was to analyze the differences between classes with and without a special emphasis regarding the level of and change in children's competence beliefs. In addition, children's mathematical thinking skills were explored in one study (III).

One of the main results arising from this study was that, even though there were initial differences between classes with and without a special emphasis in the level of children's competence beliefs and mathematical thinking skills, most of the differences were explained by background variables. These findings were interesting and contradicted the hypothesis that was the launching idea behind this thesis. A shared assumption among well-off parents in Finland has been that studying in a class with a special emphasis would have beneficial effects on children's motivation or learning results (Kosunen & Carrasco, 2016; Kosunen & Seppänen, 2015). It has been noted that classes with a special emphasis and school choices increase the within-school differences (i.e. difference between classes inside schools) in learning results, as these classes attract especially pupils with better achievement . In addition, it has been speculated that pupils studying in these selective classes would get some extra boost for their learning, which could further increase the differences between schools and classes (Kosunen, 2014; Bernelius, 2013; Kupiainen & Hotulainen, 2019). However, no studies before this one had examined whether this assumption has any empirical grounds.

The engrossing findings from this thesis showed no beneficial value-added effects of classes with a special emphasis on either pupils' competence beliefs or mathematical thinking skills, which was not entirely unexpected in light of prior international research. Results from prior studies of tracking effects (see Chapter 2.4) have been quite inconsistent and offered faltering grounds for hypothesis building. Theoretically, it seems reasonable to expect that different grouping practices would influence pupils' learning and self-beliefs, but empirical evidence has been ambiguous. Studies have shown that analyzing the effects of tracking can be a challenging task due to many confounding factors and changing research settings (e.g., Dicke et al., 2018; Hattie, 2002; 2009; Trautwein et al., 2005) that pose a challenge when attempting to compare the results from different studies. Therefore, it has been emphasized that when tracking effects are examined, studies should always clearly describe the characters of the tracking context being examined (Trautwein et al., 2005). In this thesis, I suggested that the term opt-in tracking proposed by Trautwein and colleagues (2005) could best describe the Finnish interest-based selection into classes with a special emphasis within basic

education. In the Finnish model, selection criteria for different classes with a special emphasis vary depending on the subject being emphasized (Varjo & Kalalahti, 2015, 2019). Pupils' interest and aptitudes are the most common criteria for selection (Varjo & Kalalahti, 2015, 2019) and therefore selection criteria are wider than in achievement grouping (see Trautwein et al., 2005).

Even though most analyses in studies I-III showed that pupils' competence beliefs and mathematical thinking skills developed similarly in classes with and without a special emphasis, pairwise analyses in study II showed some interesting differences in certain classes, which supported the Big-Fish-Little-Pond effect type of development. Especially in arts and language classes, the development of certain action-control beliefs was more negative than in other classes (see study II). Then again, science classes differed from other emphasized classes because there were more positive changes in some beliefs (see study II). Therefore, these results showed the need for further studies on this topic. Future studies should especially focus on analyzing the differences between different types of emphasized classes in order to detect the detailed differences between them.

Although the main interest in this thesis was to analyze the value-added effect of class type, differences in background factors between different class types were examined in more detail in one study (III). These analyses revealed some interesting gender differences between classes with and without a special emphasis and suggested that effects of emphasized classes might vary between genders. The results from study II showed that even though gender difference in academic self-concept favored boys as expected based on prior research (Marsh, 1989; Parker et al., 2018), studying in a class with a special emphasis seemed to protect girls' mathematical self-concept from a decline, which occurred in classes without a special emphasis (see study II). This finding was interesting, and it establishes a need for further studies. The gender difference in achievement has been reasonably large in Finland when compared to other OECD countries (Leino et al., 2019; OECD, 2020; Vettenranta et al., 2016). Therefore, future studies should consider the possible role that classes with a special emphasis may have in amplifying gender differences.

5.1.4 Peer effects: Big-Fish-Little-Pond, Reflected Glory and Peer Spillover effect

The findings from study II suggested that in some classes with a special emphasis the development of pupils' competence beliefs would be more negative than in other classes, therefore giving evidence of the Big-Fish-Little-Pond phenomenon. Thus, studies III and IV explored the Big-Fish-Little-Pond effect in more detail. In addition, competing peer effects, i.e., Reflected Glory- and Peer Spillover effects, were also examined in study III.

The results from studies III and IV showed that Big-Fish-Little-Pond effect was visible in the Finnish context, as expected. In other words, the results showed that even though the relation between mathematical thinking skills and mathematics self-concept was positive at the individual level, at the class level the relation became negative. Therefore, as expected by BFLPE-hypothesis, the average class-level achievement predicted an individual pupil's self-concept negatively. The result was no surprise, as BFLPE has been studied extensively and prior studies have shown it to be widely generalizable and applicable to many countries and educational contexts (Marsh et al., 2015; Marsh, Parker, et al., 2019; Seaton et al., 2009).

As study III had confirmed the existence of BFLPE in the Finnish context, the final study IV focused solely on this interesting phenomenon. Even though BFLPE has been studied extensively, most studies have analyzed it either at the school or class level and few studies have proceeded at a more local level (but see, Wouters et al., 2013). Therefore, the results from study IV offered novel information for the field of BFLPE research by testing the effect simultaneously at the level of class and smaller peer groups. The results from study IV showed that when peer group and class were examined separately, BFLPE was visible at both levels, but when both levels were scrutinized at the same time, the effect of peer group became non-significant, while the negative effect of class remained relatively unaffected. In other words, results from study IV suggest that classmates' role in the formation of self-concept would be more influential than self-chosen peers within classrooms. This finding was in line with the research by Wouters and colleagues (2013). In the first and only research before this, Wouters and colleagues (2013) examined BFLPE at the level of class and friendship groups in a cross-sectional setting and concluded that classmates' average achievement level seemed to be a more important reference frame in the formation of academic self-concept than friends' average achievement. Therefore, neither their results nor the results from this thesis (study IV) supported the local dominance theory (see, Zell & Alicke, 2010) claiming that self-evaluations would be most strongly influenced by the most proximal frame of reference (i.e., peer group instead of all classmates).

Regarding other examined peer effects (i.e., Peer Spillover and Reflected Glory effect), the results from study III did not give support for their existence. This finding was not surprising for the Peer Spillover effect, as even though some studies (e.g., Collins & Gan, 2013; Duflo et al., 2011) found evidence of positive peer effects on pupils' achievement, the results in different studies have been inconsistent overall. Several studies have also questioned these findings and claimed them to be "phantom effects" that are due to inadequate statistical methods (e.g., Dicke et al., 2018; Harker & Tymms, 2004; Marks, 2010; Nash, 2003). The findings from study III supported this view, as analyses showed that that the Peer Spillover effect disappeared when covariates (i.e., class type, mother's education level, gender) were included in the analyses (see study III).

More surprising was that the Reflected Glory effect was not supported in the analyses (see study III). Even though prior studies have shown BFLPE to be more robust then RGE, it was still a slight surprise that RGE disappeared after covariates (i.e., mother's education level and gender) were added to the analyses. Theoretically, it would have been possible that BFLPE and RGE would have been detected at the same time, balancing each other's influence (Marsh, Seaton, et al., 2008), but our results did not support this interpretation. However, it must be noted that the data that were used did not include pupils' direct comparisons of themselves or their classes to others (see, Marsh et al., 2014). Therefore, RGE was analyzed differently (see, Trautwein et al., 2006), which may explain why it was not detected. Therefore, future studies should consider this and continue the analyses of this subject.

5.2 Limitations and recommendations for future research

Even though this dissertation has many strengths, including novel research questions that were addressed using a relatively large data set, it also has several limitations that should be considered in designing future research.

The major methodological limitation of this study was that multilevel modeling was used only in the final sub-study. Even though low intraclass correlations (ICCs) (see Appendix D) [33] justified the use of single-level analyses in studies I and II, exploiting the multilevel options would have warranted more detailed and thus more interesting research questions. Therefore, future studies should consider this and continue analyzing the effects of emphasized teaching in a multilevel setting. In addition, future studies should consider more advanced techniques for taking into account the original differences between pupils studying in different types of class. As selection into classes with a special emphasis is not a random process, but in many ways connected to pupils'

[33] It has been proposed that low ICC levels (e.g., less than 5%) would justify the use of single-level analyses, but this view has also been criticized and it has been argued that multilevel models could be beneficial even when ICCs are near zero (Geiser, 2013; Hayes, 2006; Huang, 2018).

background, it is important that this selectiveness is considered in the analyses (e.g., Dicke et al., 2018; Harker & Tymms, 2004). There was an attempt to acknowledge these original differences between pupils in classes with and without a special emphasis in the analyses of the sub-studies by adding covariates to the model and exploring the effects of class type progressively with and without covariates; however, future studies should consider more effective techniques for this. For example, exploiting the technique of propensity score matching (PSM) could be useful and would allow a quasi-experimental method for comparing similar types of pupils in different types of classes (Rosenbaum & Rubin, 1983).

Another important limitation of this study was that detailed differences between different class types (i.e., classes with different emphasized subjects) were explored in only one sub-study (study II). Despite the relatively large sample, the major limitation of this study was that the number of pupils in different emphasized school subjects (e.g., mathematics, languages, music and arts) at the primary school level was too small to analyze the effects separately for each subject. Therefore, the primary school data did not enable examinations between different types of emphasized classes. However, as the results from study II showed, these focused examinations could be essential for gaining a better understanding about the possible effects of classes with a special emphasis and therefore future studies should consider this as well.

Although the major result of this study was that class type did not boost pupils' competence beliefs or mathematical thinking skills over the effects of background factors, it is possible that class type would play a role in some other facet of learning. Therefore, future studies with broader measurements of motivation and other aspects of learning (e.g., classroom climate and well-being) should be conducted. Future research should especially focus on the possible effects of studying in a class with a special emphasis on the development of achievement. In this thesis, cognitive skills (i.e., mathematical thinking skills) were explored in only one sub-study and therefore this thesis could only scratch the surface of this theme. Future studies focusing on different aspects and measures of achievement (i.e., different domains and high-stakes achievement tests) should therefore be carried out. Studies in the future should also extend the analyses to all stages of comprehensive school. In this thesis, the lower secondary school years were examined only in one sub-study (study II) and therefore more research focusing on this level should be conducted.

This research was conducted so that original studies considered either primary school (grades four to six) (Studies I, III and IV) or lower secondary school (grades seven to nine) (Study III). This framing was done because of the desire to explore the effects of class type in a setting in which the class composition remained stable over the period scrutinized. However, in the future, it would be interesting to study at least some of the themes of this dissertation in a longitudinal design with data from both primary and lower secondary school being exploited

in the same analyses. From the perspective of school choice, this research design would make it possible to study important questions regarding the transition to lower secondary school and to examine whether studying in a class with a special emphasis predicts pupils' future school choices either at the lower secondary school stage or later. In addition, a wider perspective would be interesting regarding the peer effects examined. Exploring the transition from primary to lower secondary school regarding, for example, Big-Fish-Little-Pond phenomenon would provide important information about the stability of and change in BFLPE and would make it possible to study how prior reference frames reflect an individual's self-concepts. Therefore, future studies should pursue the methodological challenge and examine the transition phase.

As the main aim in this study was to analyze the effects of certain class types in relation to the development of an individual's competence beliefs, the analyses did not include different theoretical constructs of motivation in the same analysis model. However, future studies should continue with broader analysis models and include different theoretical constructs in the same model. This would make possible to elaborate the framework presented in Chapter 2.6 and provide a more insightful understanding both about the phenomenon being examined but also about the relations between the different constructs. One interesting theory to consider in the future would be self-determination theory (Deci & Ryan, 2000). Self-determination theory with aspects of autonomy and relatedness (in addition to competence) would be well-suited to the research design presented in this study and would enable to set justified hypothesis regarding the effects of choice and certain type of social environment on individual pupil's learning.

5.3 Concluding remarks and implications for educational policy and pedagogical practices

The main goal of this study was to examine the differences between classes with and without a special emphasis and to find out whether the assumed beneficial motivational effects concerning classes with a special emphasis would have any empirical grounds. Based on findings from empirical sub-studies, it is possible to conclude that although classes with a special emphasis clearly differed from classes without an emphasis regarding a pupil's background, these differences were not reflected in the development of either competence beliefs or mathematical thinking skills. Almost all differences between classes with and without a special emphasis disappeared when differences in background variables were considered. The development in different types of class was strikingly similar. Therefore, the results of this thesis did not support the assumed beneficial effects of emphasized teaching.

More importantly, the analyses showed that studying in a class without a special emphasis did not by any means seem to be less favorable in terms of the

development of action-control beliefs (see studies I and II) or academic self-concept or mathematical thinking skills (see study III). On the contrary, the development of some specific beliefs in classes without a special emphasis was even more positive than in some classes with a special emphasis (see study II) and none of the detected signs of negative development (see study II) occurred in classes without a special emphasis. Similarly, results from study III showed that BFLPE overrode other positive peer effects (i.e., RGE and PSE), indicating that studying in a high-achieving group may negatively influence an individual's self-concept.

To put it simply, the results of this thesis showed that, for the most, the development of competence beliefs and thinking skills was similar in different types of classes. However, results from studies II and III showed that, contrary to the beneficial effects that were expected, studying in a selective peer group in a class with a special emphasis seemed to have detrimental consequences for some pupils' self-perceptions. Therefore, the findings of this thesis should be considered when the meaningfulness of the current school choice policy on classes with a special emphasis is evaluated. Also considering findings regarding the equality effects of tracking (e.g., Hanushek & Wößmann, 2006; Hattie, 2002) and classes with a special emphasis (Kosunen, Bernelius, et al., 2016; Seppänen et al., 2012; Seppänen, Kalalahti, et al., 2015), the findings of this thesis, showing negative changes in pupils' competence beliefs, challenge the politics and the officials responsible for education to reconsider the rationale behind current practices of organizing teaching with a special emphasis. One option could be to organize teaching of selective subjects within general classes and to mix pupils into the same classes. Then, pupils choosing teaching with a special emphasis would have more lessons in the those subjects but would otherwise study in same classroom as other pupils. This would maintain the original idea of individuality behind classes with a special emphasis, but at the same time, conserve the idea of comprehensive school and equal educational opportunities. At the same time, "mixed sorting" could reduce the negative effects of selective peer groups. These "mixed practices" have already been used in some schools or municipalities (Simola et al., 2015), but the main practice has been to allocate pupils in emphasized teaching into their own teaching groups i.e. classes with a special emphasis. However, findings from this thesis clearly show the importance of further investigating and reconsidering these practices.

Overall, the results of this thesis shed light on the possible consequences that certain education policies may have on individual pupils' learning. Even though the results of this thesis cannot be generalized to educational contexts other than the Finnish one as such, the results clearly show that the way pupils are sorted into groups may have consequences on individual pupils' learning. The classroom is an important social environment in which a child develops and interacts with peers and teachers (Urdan & Schoenfelder, 2006) and the way this environment is

composed is an educational political question that can be reflected in many ways on individuals' learning. Equality has been one of the core values in Finnish education policy. The central idea behind the current comprehensive school system has been the equality of educational opportunities. It has been seen as important that common comprehensive school brings together children from different backgrounds and offers them all the same possibilities to learn (Ahonen, 2001). However, as prior studies and the results of this thesis have shown, classes with a special emphasis clearly break this emphasis of equality by sorting pupils from higher socio-economic families with better school achievements into separate classes. Thus, it is evident that classes with a special emphasis differentiate the schooling experiences of children and change the composition of learning environments. According to the surprising results of this thesis, instead of the assumed beneficial effects, the only detected effects of classes with a special emphasis were negative. Therefore, these findings both illustrate the difficulty of forecasting the implications of certain policy as well as highlight the importance of study them.

References

Ahonen, S. (2001). Kuka tarvitsee yhtenäistä koulua? [Who needs the comprehensive school?]. In A. Jauhiainen, R. Rinne & J. Tähtinen (Eds.), *Koulutuspolitiikka Suomessa ja ylikansalliset mallit.* [*Educational Policy in Finland and Global Model*] (pp. 155–184). Suomen kasvatustieteellinen seura. Turku: Painosalama Oy. ISBN: 952-5401-00-6

Albert, S. (1977). Temporal comparison theory. *Psychological Review, 84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Altermatt, E. R., & Pomerantz, E. M. (2005). The Implications of Having High-achieving Versus Low-achieving Friends: A Longitudinal Analysis. *Social Development*, *14*(1), 61–81. https://doi.org/10.1111/j.1467-9507.2005.00291.x

Ball, S. J., Bowe, R., & Gewirtz, S. (1995). Circuits of Schooling: A Sociological Exploration of Parental Choice of School in Social Class Contexts. *The Sociological Review, 43*(1), 52–78. https://doi.org/10.1111/j.1467-954X.1995.tb02478.x

Ball, S. J., Bowe, R., & Gewirtz, S. (1996). School choice, social class and distinction: the realization of social advantage in education. *Journal of education policy*, 11(1), 89-112. https://doi.org/10.1080/0268093960110105

Ball, S. J., & Vincent, C. (1998). 'I Heard It on the Grapevine': 'Hot' knowledge and school choice. *British Journal of Sociology of Education*, *19*(3), 377–400. https://doi.org/10.1080/0142569980190307

Ball, S. J., Vincent, C., Kemp, S., & Pietikainen, S. (2004). Middle Class Fractions, Childcare and the 'Relational' and 'Normative' Aspects of Class Practices. *The Sociological Review*, *52*(4), 478–502. https://doi.org/10.1111/j.1467-954X.2004.00492.x

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, *84*, 191–215.

Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, *28*(2), 117–148. https://doi.org/10.1207/s15326985ep2802_3

Bandura, A. (1999). A social cognitive theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality* (pp. 154–196). New York: Guilford Publications.

Bandura, A. (2004). Swimming against the mainstream: The early years from chilly tributary to transformative mainstream. *Behaviour Research and Therapy*, *42*(6), 613–630. https://doi.org/10.1016/j.brat.2004.02.001

Belfi, B., Haelermans, C., & De Fraine, B. (2016). The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach. *British Journal of Educational Psychology*, *86*(4), 501–525. https://doi.org/10.1111/bjep.12120

Berisha, A.-K., & Seppänen, P. (2017). Pupil selection segments urban comprehensive schooling in Finland: Composition of school classes in pupils' school performance, gender, and ethnicity. *Scandinavian Journal of Educational Research*, *61*(2), 240–254. https://doi.org/10.1080/00313831.2015.1120235

Bernelius, V. (2013). Eriytyvät kaupunkikoulut: Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä. *[Differentiation of urban schools. The differences of pool of pupils, parental school choice and learning outcomes in connection with neighbourhood effects as part of the segregation of the city of Helsinki]*. Helsinki: Helsingin kaupungin tietokeskus. ISBN: 978-952-272-529-5

Betts, J. R. (2011). The Economics of Tracking in Education. In E. Hanushek, S. Machin & L. Wößmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 341–381). Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.00007-7

Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology, 76*(3), 420–430. https://doi.org/10.1037/0022-3514.76.3.420

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' Experiences of Ability Grouping—Disaffection, polarisation and the construction of failure. *British Educational Research Journal*, *26*(5), 631–648. https://doi.org/10.1080/713651583

Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology*, 41(4), 377-397. https:/ /doi/10.1111/j.1464-0597.1992.tb00713.x

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley. http://doi.org/ 10.1002/9781118619179

Bong, M., & Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review*, *15*(1), 1–40. https://doi/10.1023/A:1021302408382

Bourdieu, P. (1989). Social Space and Symbolic Power. *Sociological Theory*, 7(1), 14–25. https://www.jstor.org/stable/202060

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, 32(7), 513. https://doi/10.1037/0003-066X.32.7.513

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American psychologist*, 34(10), 844. https://doi/10.1037/0003-066X.34.10.844

Bronfenbrenner, U. (1992). Ecological Systems Theory. Reprinted In U. Bronfenbrenner (Eds.), *Making human beings human: Bioecological perspectives on human development*. California: Sage. ISBN: 0-7619-271-3

Bronfenbrenner, U. (2001). The Bioecological Theory of Human Development. Reprinted In U. Bronfenbrenner (Eds.), *Making human beings human: Bioecological perspectives on human development*. California: Sage. ISBN: 0-7619-271-3

Bronfenbrenner, U. (2005). Making human beings human: Bioecological perspectives on human development. California: Sage. ISBN: 0-7619-271-3

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*(52), 782–861. https://doi.org/10.1111/j.1468-0327.2007.00189.x

Burke, M. A., & Sass, T. R. (2013). Classroom Peer Effects and Student Achievement. *Journal of Labor Economics*, *31*(1), 51–82. https://doi.org/10.1086/666653

Burns, R. B., & Mason, D. A. (2002). Class Composition and Student Achievement in Elementary Schools. *American Educational Research Journal*, *39*(1), 207–233. https://doi.org/10.3102/00028312039001207

Butler, R., & Hasenfratz, L. (2017). Gender and competence motivation. In A. Elliot, C. Dweck & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (Second edition, pp. 489–511). New York: The Guilford Press. ISBN: 978-1-4625-3603-0

Carman, K. G., & Zhang, L. (2012). Classroom peer effects and academic achievement: Evidence from a Chinese middle school. *China Economic Review*, *23*(2), 223–237. https://doi.org/10.1016/j.chieco.2011.10.004

Cham, H., Reshetnyak, E., Rosenfeld, B., & Breitbart, W. (2017). Full Information Maximum Likelihood Estimation for Latent Variable Interactions With Incomplete Indicators. *Multivariate Behavioral Research*, *52*(1), 12–30. https://doi.org/10.1080/00273171.2016.1245600

Chang, R., Adams, N. & Little, T.D. Action-Control Beliefs and Agentic Actions. In M.L. Wehmeyer, K.A. Shogren, T.D. Little. S. J. Lopez (Eds.), *Development of Self-Determination through the Life-Course*. https://doi.org/10.1007/978-94-024-1042-6_22

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiu, M. M., Chow, B. W.-Y., & Joh, S. W. (2017). Streaming, tracking and reading achievement: A multilevel analysis of students in 40 countries. *Journal of Educational Psychology*, *109*(7), 915–934. https://doi.org/10.1037/edu0000188

Cimpian, A. (2017). Early reasoning about competence is not irrationally optimistic, nor does it stem from inadequate cognitive representations. In A. Elliot, C. Dweck & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (Second edition, pp. 387–407). New York: The Guilford Press. ISBN: 978-1-4625-3603-0

Coleman, J. M., & Fults, B. (1985). Special-Class Placement, Level of Intelligence, and the Self-Concepts of Gifted Children: A Social Comparison Perspective. *Remedial and Special Education*, *6*(1), 7–12. https://doi.org/10.1177/074193258500600102

Collins, C., & Gan, L. (2013). Does sorting students improve scores? An analysis of class composition. (No. W18848). *National Bureau of Economic Resear*ch. http://www.nber.org/papers/w18848

Cooc, N., & Kim, J. S. (2017). Peer influence on children's reading skills: A social network analysis of elementary school classrooms. *Journal of Educational Psychology*, *109*(5), 727–740. https://doi.org/10.1037/edu0000166

Dai, D. Y., & Rinn, A. N. (2008). The Big-Fish-Little-Pond Effect: What Do We Know and Where Do We Go from Here? *Educational Psychology Review*, *20*(3), 283–317. https://doi.org/10.1007/s10648-008-9071-x

Deci, E.L. & Ryan, M. R. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227-268.

De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal*, *29*(6), 841–859. https://doi.org/10.1080/0141192032000137330

Demetriou, A., & Kazi, S. (2006). Self-awareness in g (with processing efficiency and reasoning). *Intelligence*, *34*(3), 297–317. https://doi.org/10.1016/j.intell.2005.10.002

Demetriou, A., Kyriakides, L., & Avraamidou, C. (2003). The missing link in the relations between intelligence and personality. *Journal of Research in Personality*, *37*(6), 547–581. https://doi.org/10.1016/S0092-6566(03)00048-5

Demetriou, A., Pachaury, A., Metallidou, Y., & Kazi, S. (1996). Universals and Specificities in the Structure and of Quantitative-relational Thought: A Cross-cultural Study in Greece and India. *International Journal of Behavioral Development*, *19*(2), 255–290. https://doi.org/10.1080/016502596385785

Demetriou, A., Platsidou, M., Efklides, A., Metallidou, Y., & Shayer, M. (1991). The development of quantitative-relational abilities from childhood to adolescence: Structure, scaling, and individual differences. *Learning and Instruction*, *1*(1), 19–43. https://doi.org/10.1016/0959-4752(91)90017-3

Demetriou, A., Spanoudis, G., & Mouyi, A. (2011). Educating the Developing Mind: Towards an Overarching Paradigm. *Educational Psychology Review*, *23*(4), 601–663. https://doi.org/10.1007/s10648-011-9178-3

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*, *110*(8), 1112–1126. https://doi.org/10.1037/edu0000259

Diener, E., & Fujita, F. (1997). Social comparisons and subjective well-being. In B. Buunk, F. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–357). Lawrence Erlbaum Associates Publishers.

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social Comparison in the Classroom: A Review. *Review of Educational Research*, *78*(4), 828–879. https://doi.org/10.3102/0034654308321210

Diseth, Å., Meland, E., & Breidablik, H. J. (2014). Self-beliefs among students: Grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learning and Individual Differences*, *35*, 1–8. https://doi.org/10.1016/j.lindif.2014.06.003

Dracup, T. (2014). The politics of setting. Retrieved [7.5.2021] from: https://giftedphoenix.wordpress.com/2014/11/12/the-politics-of-setting/.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review*, *101*(5), 1739–1774. https://doi.org/10.1257/aer.101.5.1739

Dupriez, V., Dumay, X., & Vause, A. (2008). How Do School Systems Manage Pupils' Heterogeneity? *Comparative Education Review*, *52*(2), 245–273. https://doi.org/10.1086/528764

Duru-Bellat, M., & Mingat, A. (1998). Importance of Ability Grouping in French "Collèges" and its Impact upon Pupils' Academic Achievement. *Educational Research and Evaluation*, *4*(4), 348–368. https://doi.org/10.1076/edre.4.4.348.6951

Dweck, C. S. (1986). Motivational Processes Affecting Learning. *American Psychologist*, 1040–1048. https://doi/10.1037/0003-066X.41.10.1040

Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioral Development*, *35*(3), 195–201. https://doi.org/10.1177/0165025411398185

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed.In W. Damon & N. Eisenberg (Eds.), *Handbook of Child Psychology. Social, emotional, and personality development.* (Fifth Edition, pp. 1017-1096). John Wiley & Sons. ISBN: 0-471-34981

Eccles, & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development*, *64*(3), 830. https://doi.org/10.2307/1131221

Elliot, A. & Hulleman, C. (2017). Achievement goals. In A. Elliot, C. Dweck & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (Second edition, pp. 43–60). New York: The Guilford Press. ISBN: 978-1-4625-3603-0

Epple, D., Newlon, E., & Romano, R. (2000). *Ability Tracking, School Competition, and the Distribution of Educational Benefits* (No. w7854; p. w7854). National Bureau of Economic Research. https://doi.org/10.3386/w7854

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The Big-Fish-Little-Pond Effect on Academic Self-Concept: A Meta-Analysis. *Frontiers in Psychology*, *9*, 1569. https://doi.org/10.3389/fpsyg.2018.01569

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th edition). London: SAGE Publications. ISBN: 978-1-5264-1952-1

Finch, W. H., & Bolin, J. E. (2017). *Multilevel modeling using Mplus*. Boca Raton: Chapman & Hall/CRC. ISBN: 978-1-4987-4824-7

Francis, B., Connolly, P., Archer, L., Hodgen, J., Mazenod, A., Pepper, D. & Travers, M. C. (2017). Attainment Grouping as self-fulfilling prophesy? A mixed methods exploration of self confidence and set level among Year 7 students. *International Journal of Educational Research*, 86, 96-108. https://doi.org/10.1016/j.ijer.2017.09.001

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of Mathematics Interest in Adolescence: Influences of Gender, Family, and School Context: Development of mathematics interest in adolescence. *Journal of Research on Adolescence*, *20*(2), 507–537. https://doi.org/10.1111/j.1532-7795.2010.00645.x

Geiser, C. (2013). *Data analysis with Mplus*. New York: The Guilford Press. ISBN: 978-1-4625-0-2455

Geldhof, J., & Little, T. D. (2011). Influences of Children's and Adolescents' Action-Control Processes on School Achievement, Peer Relationships, and Coping with Challenging Life Events. In R. M. Lerner, J. V. Lerner, E. P. Bowers, S. Lewin-Bizan, S. Gestsdottir, & J. B. Urban (Eds.), *Thriving in childhood and adolescence: The role of self-regulation processes. New Directions for Child and Adolescent Development.* (pp. 45–59). https://doi.org/10.1002/cd.303

Gewirtz, S., Ball, S. J., & Bowe, R. (1995). *Markets, choice, and equity in education*. Buckingham: Open University Press. ISBN-13: 978-0-3351-9369-1

Gottfried, M. A. (2014). Peer Effects in Urban Schools: Assessing the Impact of Classroom Composition on Student Achievement. *Educational Policy*, *28*(5), 607–647. https://doi.org/10.1177/0895904812467082

Guay, F., Larose, S., & Boivin, M. (2004). Academic Self-concept and Educational Attainment Level: A Ten-year Longitudinal Study. *Self and Identity*, *3*(1), 53–68. https://doi.org/10.1080/13576500342000040

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles*, *66*(3–4), 153–166. https://doi.org/10.1007/s11199-011-9996-2

Guo, J., Marsh, H. W., Parker, P. D., & Dicke, T. (2018). Cross-cultural generalizability of social and dimensional comparison effects on reading, math, and science self-concepts for primary school students using the combined PIRLS and TIMSS data. *Learning and Instruction*, *58*, 210–219. https://doi.org/10.1016/j.learninstruc.2018.07.007

Hanushek, E. A., & Wößmann, L. (2006). Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries. *The Economic Journal*, *116*(510), C63–C76. https://doi.org/10.1111/j.1468-0297.2006.01076.x

Harker, R., & Tymms, P. (2004). The Effects of Student Composition on School Outcomes. *School Effectiveness and School Improvement*, *15*(2), 177–199. https://doi.org/10.1076/sesi.15.2.177.30432

Hattie. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, *37*(5), 449–481. https://doi.org/10.1016/S0883-0355(03)00015-6

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Reprinted). Routledge. ISBN: 978-0-4154-7618-8

Hautamäki, Arinen, P., Hautamäki, A., Kupiainen, S., Lindblom, B., Niemivirta, M., Pakaslahti, L., Rantanen, P., & Scheinin, P. (2002). *Assessing learningto- learn. A framework.* Helsinki: National Board of Education. ISBN: 952-13-1423-0

Hautamäki, Arinen, P., Hautamäki, A., Lehto, A., Lindblom, B., Kupiainen, S., Outinen, K., Pekuri, M., Reuhkala, M., & Scheinin, P. (2001). *Ensiaskeleet – oppimisen edellytykset: Luokanopettajille tarkoitetun seulan toimivuus Helsinki-aineiston perusteella. [First steps – Prerequisities of learning: The functionality of the screening test for class teachers based on Helsinki-data].* Publications of Helsinki Education Department, vol. 2001 (No.A17). Helsinki: City of Helsinki, Education Depatment. ISBN: 951-718-738-6

Hautamäki, Hautamäki, A., & Kupiainen, S. (2010). Assessment in Schools – Learning to Learn. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education.* (3ed. Vol3. pp. 268–272). Oxford: Elsevier. ISBN: 978-0-08-044897-8

Hautamäki, & Kupiainen, S. (2014). Learning to learn in Finland: Theory and policy, research and practice. In R. Deakin Crick, C. Stringher & K. Ren (Eds.), *Learning to Learn: International perspectives from theory and practice.* (pp. 179–205). http://doi.org/10.4324/9780203078044-9

Hayes, A. F. (2006). A Primer on Multilevel Modeling. *Human Communication Research*, *32*(4), 385–410. https://doi.org/10.1111/j.1468-2958.2006.00281.x

Heiskala, R., & Luhtakallio, (2006). Johdanto: Suunnittelutaloudesta kilpailukyky-yhteiskuntaan. [From Planning Society to Competition Society]. In R. Heiskala & E. Luhtakallio (Eds.), *Uusi jako: Miten Suomesta tuli kilpailukyky-yhteiskunta.* Helsinki: Gaudeamus.

Hienonen, N., Lintuvuori, M., Jahnukainen, M., Hotulainen, R., & Vainikainen, M.-P. (2018). The effect of class composition on cross-curricular competences – Students with special educational needs in regular classes in lower secondary education. *Learning and Instruction, 58,* 80–87. https://doi.org/10.1016/j.learninstruc.2018.05.005

Hienonen, N. (2020). *Does class placement matter? Students with special educational needs in regular and special classes.* [Doctoral dissertation]. Helsinki Studies in Education, number 87. University of Helsinki: Unigrafia. ISBN: 978-951-51-6392-9

Horn, D. (2009). Age of selection counts: A cross-country analysis of educational institutions. *Educational Research and Evaluation*, *15*(4), 343–366. https://doi.org/10.1080/13803610903087011

Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2015). Does classroom composition make a difference: Effects on developments in motivation, sense of classroom belonging, and achievement in upper primary school. *School Effectiveness and School Improvement*, *26*(2), 125–152. https://doi.org/10.1080/09243453.2014.887024

Hosenfeld, B., Boom, D. C., & Resing, W. C. M. (1997). Constructing Geometric Analogies for the Longitudinal Testing of Elementary School Children. *Journal of Educational Measurement*, *34*(4), 367–372. https://doi.org/10.1111/j.1745-3984.1997.tb00524.x

Hoxby, C. (2000). Peer effects in the classroom. Learning from gender and race variation. (No. W7867). In *National Bureau of Economic Research.* Cambridge. http://doi.org/10.3386/w7867

Huang, F. L. (2018). Multilevel modeling myths. *School Psychology Quarterly*, *33*(3), 492–499. https://doi.org/10.1037/spq0000272

Huguet, P., Dumas, F., Marsh, H., Régner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish–little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology*, *97*(1), 156–170. https://doi.org/10.1037/a0015558

Huguet, P., Dumas, F., Monteil, J. M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, *31*(5), 557–578. https://doi.org/10.1002/ejsp.81

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Ireson, J., Hallam, S., & Plewis, I. (2001). Ability grouping in secondary schools: Effects on pupils' self-concepts. *British Journal of Educational Psychology*, *71*(2), 315–326. https://doi.org/10.1348/000709901158541

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in Children's Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development*, *73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

James, W. (1890). *The principles of psychology* (Reprint, Vols. 1 & 2 Complete and Unabridged with Illustrations and Tables). Pantianos Classics. ISBN: 978-1543183184

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond* (Third Edition). Routledge, Taylor & Francis Group. ISBN: 978-1-138-81983-2

Kalalahti, M., Silvennoinen, H., & Varjo, J. (2015). Kouluvalinnat kykyjen mukaan? – Erot painotettuun opetukseen valikoitumisessa. *Kasvatus*, *46*(1), 19–35.

Kalalahti, M., Silvennoinen, H., Varjo, J., & Rinne, R. (2016). Education for all? Urban Finnish parental attitudes towards the universalism and selectivism of comprehensive school system. In P. Seppänen, A. Carrasco, M. Kalalahti, R. Rinne & H. Simola (Eds.), *Contrasting Dynamics in Education Politics of Extremes. School Choice in Chile and Finland.* (Publication series: Comparative and International Education: A Diversity of Voices. Vol. 37, pp. 205–224). Rotterdam: Sense Publisher. ISBN: 978-94-6300-260-8

Kalalahti, M., & Varjo, J. (2012). Tasa-arvo ja oikeudenmukaisuus perusopetukseen sijoittumisessa ja valikoitumisessa. *Kasvatus & Aika*, *6*(1), 39–55.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, *61*(3), 458–495. https://doi.org/10.1016/j.jue.2006.07.006

Kelly, A. (2001). *Do Increased Levels of Parental Involvement Account for Social Class Differences in Track Placement?* [Research Reports 143]. https://files.eric.ed.gov/fulltext/ED451447.pdf

Kindermann, T. A. (2007). Effects of Naturally Existing Peer Groups on Changes in Academic Engagement in a Cohort of Sixth Graders. *Child Development*, *78*(4), 1186–1203. https://doi.org/10.1111/j.1467-8624.2007.01060.x

Kindermann, T. A. (2016). Peer Group Inlfluences on Students' Academic Motivation. In K. Wentzel & G. Ramani (Eds.), *Handbook of Social Influences in School Contexts Social-Emotional, Motivation and Cognitive Outcomes* (pp. 31–46). Routledge. ISBN-13: 978-1138781405

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed). New York: Guilford Press. ISBN: 978-1-57230-690-5

Kosunen, S. (2014). Reputation and parental logics of action in local school choice space in Finland. *Journal of Education Policy*, *29*(4), 443–466. https://doi.org/10.1080/02680939.2013.844859

Kosunen, S., Bernelius, V., Seppänen, P., & Porkka, M. (2016). School Choice to Lower Secondary Schools and Mechanisms of Segregation in Urban Finland. *Urban Education*, 004208591666693. https://doi.org/10.1177/0042085916666933

Kosunen, S., Bernelius, V., Seppänen, P., & Porkka, M. (2020). School choice to lower secondary schools and mechanisms of segregation in urban Finland. *Urban education*, *55*(10), 1461-1488. https://doi.org/10.1177%2F0042085916666933

Kosunen, S., & Carrasco, A. (2016). Parental preferences in school choice: Comparing reputational hierarchies of schools in Chile and Finland. *Compare: A Journal of Comparative and International Education*, *46*(2), 172–193. https://doi.org/10.1080/03057925.2013.861700

Kosunen, S., & Seppänen, P. (2015). The transmission of capital and a feel for the game: Upper-class school choice in Finland. *Acta Sociologica*, *58*(4), 329–342. https://doi.org/10.1177/0001699315607968

Kosunen, S., Seppänen, P., & Bernelius, V. (2016). Naapurustojen segregaatio ja kaupunkilaisperheiden eriytyvät kouluvalintastrategiat. *Kasvatus*, *47*(3), 230–244.

Kraska, M. (2010). Repeated Measures Design. In *Encyclopedia of Research Design*. SAGE Publications, Inc. https://doi.org/10.4135/9781412961288

Kuosmanen, I. 2020. Essays on Economics of Immigration and Education. Doctoral dissertation. University of Helsinki, Faculty of Social Sciences. ublication series of the Helsinki Center of Economic Research - URN:ISSN:2323-9786

Kupiainen, S., & Hotulainen, R. (2019). Erilaisia luokkia, erilaisia oppilaita. In J. Hautamäki, I. Rämä, & M.-P. Vainikainen (Eds.), *Perusopetus, tasa-arvo ja oppimaan oppiminen: Valtakunnallinen arviointitutkimus peruskoulun päättövaiheesta* (pp. 139–165). Kasvatustieteellisiä tutkimuksia 52. Helsinki: Helsingin yliopisto.

Lauder, H. (Ed.). (1999). *Trading in futures: Why markets in education don't work*. Buckingham: Open University Press. ISBN: 978-0335202775

Leaper, C. (2013). Gender Development During Childhood. In P.D. Zelazo (Eds.), *The Oxford Handbook of Developmental Psychology: Vol 2: Self and Other*. https://doi.org/10.1093/oxfordhb/9780199958474.013.0014

Lei, P.-W., & Wu, Q. (2007). Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement: Issues and Practice*, *26*(3), 33–43. https://doi.org/10.1111/j.1745-3992.2007.00099.x

Leino, K., Ahonen, A., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., Lämsä, J., Nissinen, K., Nissinen, V., Puhakka, E., Pulkkinen, J., Rautopuro, J., Siren, M., Vainikainen, M.-P., & Vettenranta, J. (2019). *PISA 18 ensituloksia*. *Suomi parhaiden joukossa. Suomalaisnuorten lukutaidon ja lukuharrastuksen muuttuminen.* Opetus- ja kulttuuriministeriön julkaisuja 2019:40. Helsinki: Opetus- ja kulttuuriministeriö.

Little. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Little, T.D. (2002). Agency in development. In W.H. Hartup & R.K. Silbereisen (Eds.), Growing points in developmental science: An introduction. pp. (223-240). East Sussex. England: Psychology Press.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: The Guilford Press. ISBN-978-1-4625-1016-0

Little, T.D. Hawley, P., Henrich, C., & Marsland, K. (2002). Three views of the agentic self: A Developmental Synthesis. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 177–195). University of Rochester Press.

Little, T.D. & Lopez, D. (1997). Regularities in the Development of Children's Causality Beliefs About School Performance Across Six Sociocultural Contexts. *Developmental Psychology*, *33*(1), 165–175.

Little, T.D., Lopez, D. F., Oettingen, G., & Baltes, P. B. (2001). A comparative-longitudinal study of action-control beliefs and school performance: On the role of context. *International Journal of Behavioral Development*, *25*(3), 237–245. https://doi.org/10.1080/01650250042000258

Little, T.D., Stetsenko, A., & Maier, H. (1999). Action-control Beliefs and School Performance: A Longitudinal Study of Moscow Children and Adolescents. *International Journal of Behavioral Development*, *23*(3), 799–823. https://doi.org/10.1080/016502599383801

Little, T. D., Oettingen, G., Stetsenko, A., & Baltes, P. B. (1995). Children's action-control beliefs about school performance: How do American children compare with German and Russian children? *Journal of Personality and Social Psychology*, *69*(4), 686–700. https://doi.org/10.1037/0022-3514.69.4.686

Liu, Y., & Yussen, S. R. (2005). A comparison of perceived control beliefs between Chinese and American students. *International Journal of Behavioral Development*, *29*(1), 14–23. https://doi.org/10.1080/01650250444000298

Logie, R., & Pearson, D. G. (1997). The Inner Eye and the Inner Scribe of Visuo-spatial Working Memory: Evidence from Developmental Fractionation. *European Journal of Cognitive Psychology*, *9*(3), 241–257. https://doi.org/10.1080/713752559

Lönnqvist, J.-E., Verkasalo, M., & Vainikainen, M.-P. (2011). Parent-teacher agreement on 7-Year-old children's personality. *European Journal of Personality*, *25*(5), 306–316. https://doi.org/10.1002/per.791

Lopez, D. F. (1999). Social cognitive influences on self-regulated learning: The impact of action-control beliefs and academic goals on achievement-related outcomes. *Learning and Individual Differences*, *11*(3), 301–319. https://doi.org/10.1016/S1041-6080(99)80005-3

Loveless, T. (2009). Tracking and Detracking: *High achievers in Massachusetts Middle Schools.* Retrieved [26.5.2021]from: https://files.eric.ed.gov/fulltext/ED507543.pdf

Loyalka, P., Zakharov, A., & Kuzmina, Y. (2018). Catching the Big Fish in the Little Pond Effect: Evidence from 33 Countries and Regions. *Comparative Education Review*, *62*(4), 542–564. https://doi.org/10.1086/699672

Maaz, K., Trautwein, U., Ldtke, O., & Baumert, J. (2008). Educational Transitions and Differential Learning Environments: How Explicit Between-School Tracking Contributes to Social Inequality in Educational Outcomes. *Child Development Perspectives*, *2*(2), 99–106. https://doi.org/10.1111/j.1750-8606.2008.00048.x

Malmberg, L.-E., & Little, T. D. (2007). Profiles of ability, effort, and difficulty: Relationships with worldviews, motivation and adjustment. *Learning and Instruction*, *17*(6), 739–754. https://doi.org/10.1016/j.learninstruc.2007.09.014

Malmberg, L.-E., Wanner, B., & Little, T. D. (2008). Age and school-type differences in children's beliefs about school performance. *International Journal of Behavioral Development*, *32*(6), 531–541. https://doi.org/10.1177/0165025408095558

Marks, G. N. (2010). What aspects of schooling are important? School effects on tertiary entrance performance. *School Effectiveness and School Improvement*, *21*(3), 267–287. https://doi.org/10.1080/09243451003694364

Marsh, H. W. (1984). Self-concept: The Application of a Frame of Reference Model to Explain Paradoxical Results. *Australian Journal of Education*, *28*(2), 165–181. https://doi.org/10.1177/000494418402800207

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

Marsh, H. W. (1989). Age and Sex Effects in Multiple Dimensions of Self-Concept: Preadolescence to Early Adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi/10.1037/0022-0663.81.3.417

Marsh, H. W., Abduljabbar, A. S., Morin, A. J. S., Parker, P., Abdelfattah, F., Nagengast, B., & Abu-Hilal, M. M. (2015). The big-fish-little-pond effect: Generalizability of social comparison processes over two age cohorts from Western, Asian, and Middle Eastern Islamic countries. *Journal of Educational Psychology*, *107*(1), 258–271. https://doi.org/10.1037/a0037485

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A Multifaceted Academic Self-Concept: Its Hierarchical Structure and Its Relation to Academic Achievement. *Journal of Educational Psychology*, *82*(4), 623–636. https://doi/10.1037/0022-0663.80.3.366

Marsh, H. W., & Craven, R. G. (2006). Reciprocal Effects of Self-Concept and Performance From a Multidimensional Perspective: Beyond Seductive Pleasure and Unidimensional Perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K.-T. (2003). Big-Fish-Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. https://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., Hau, K.-T., & Craven, R. (2004). The Big-Fish-Little-Pond Effect Stands Up to Scrutiny. *American Psychologist*, *59*(4), 269–271. https://doi.org/10.1037/0003-066X.59.4.269

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal Multilevel Models of the Big-Fish-Little-Pond Effect on Academic Self-Concept: Counterbalancing Contrast and Reflected-Glory Effects in Hong Kong Schools. *Journal of Personality and Social Psychology*, 78(2), 337–349. https://doi/10.1037/0022-3514.78.2.337

Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66. https://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-Latent Models of School Contextual Effects: Integrating Multilevel and Structural Equation Approaches to Control Measurement and Sampling Error. *Multivariate Behavioral Research*, *44*(6), 764–802. https://doi.org/10.1080/00273170903333665

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering: Academic self-concept. *British Journal of Educational Psychology*, *81*(1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Martin, A. J., Yeung, A., & Craven, R. G. (2017). Competence self perceptions. In A. Elliot, C. Dweck & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (Second edition, pp. 85–115). New York: The Guilford Press. ISBN: 978-1-4625-3603-0

Marsh, H. W., & O'Mara, A. (2008). Reciprocal Effects Between Academic Self-Concept, Self-Esteem, Achievement, and Attainment Over Seven Adolescent Years: Unidimensional and Multidimensional Perspectives of Self-Concept. *Personality and Social Psychology Bulletin*, *34*(4), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213

Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three Paradoxical Effects on Academic Self-Concept Across Countries, Schools, and Students: Frame-of-Reference as a Unifying Theoretical Explanation. *European Psychologist*, *24*(3), 231–242. https://doi.org/10.1027/1016-9040/a000332

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*(2), 263–280. https://doi.org/10.1037/dev0000393

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., & Seaton, M. (2015). The Big-Fish–Little-Pond Effect, Competence Self-perceptions, and Relativity. In *Advances in Motivation Science* (Vol. 2, pp. 127–184). Elsevier. https://doi.org/10.1016/bs.adms.2015.05.002

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-fish–little-pond-effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research. *Educational Psychology Review*, *20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J., & Köller, O. (2007). The Big-Fish-Little-Pond Effect: Persistent Negative Effects of Selective High Schools on Self-Concept After Graduation. *American Educational Research Journal*, *44*(3), 631–669. https://doi.org/10.3102/0002831207306728

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, *100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., & Yeung, A. S. (1997). Coursework Selection: Relations to Academic Self-Concept and Achievement. *American Educational Research Journal*, *34*(4), 691–720. https://doi.org/10.3102/00028312034004691

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology*, *44*(5), 351–373. https://doi.org/10.1016/j.jsp.2006.04.004

Millsapp, R., & Cham, H. (2013). Investigating factorial invariance in longitudinal data. In B. Laursen, T.D. Little & N.A. Card (Eds.), *Handbook of developmental research methods* (pp. 109–128). New York: The Guilford Press. ISBN: 978-1-4-6251-3932

Ministry of Education and Culture. (2002). The Finnish Education System. Retrieved [26.5.2021] from: https://minedu.fi/en/education-system

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

Muenks, K., Wigfield, A., & Eccles, J. S. (2018). I can do this! The development and calibration of children's expectations for success and competence beliefs. *Developmental Review*, *48*, 24–39. https://doi.org/10.1016/j.dr.2018.04.001

Müller-Kalthoff, H., Helm, F., & Möller, J. (2017). The big three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Social Psychology of Education*, *20*(4), 849–873. https://doi.org/10.1007/s11218-017-9395-9

Murrar, S., & Brauer, M. (2018). Mixed Model Analysis of Variance. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n436

Muthén, L., & Muthén, B. (2018). *Mplus: Statistical analysis with latent variables: User's guide [version 8]*. Muthén & Muthén. Retrieved [26.5.2021] from: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

Nash, R. (2003). Is the School Composition Effect Real?: A Discussion With Evidence From the UK PISA Data. *School Effectiveness and School Improvement*, *14*(4), 441–457. https://doi.org/10.1076/sesi.14.4.441.17153

Nicholls, J. G. (1978). The Development of the Concepts of Effort and Ability, Perception of Academic Attainment, and the Understanding That Difficult Tasks Require More Ability. *Child Development*, *49*(3), 800. https://doi.org/10.2307/1128250

Niemivirta, M. (2000). "Ehkä osaisinkin, mutta kun ei huvita"–Motivaatio ja koulumenestys yläasteen päättyessä. ["Maybe I could, but I don't care to"–Motivation and school achievement at the end of lower secondary education]. In J. Hautamäki, P. Arinen, A. Hautamäki, M. Ikonen-Varila, S. Kupiainen, B. Lindblom, M. Niemivirta, P. Rantanen, M. Ruuth & P. Scheinin (Eds.), Oppimaan oppiminen yläasteella. Oppimistulosten arviointi 7/2000. Helsinki: Opetushallitus.

Nordstokke, D., & Zumbo, B. (2010). A new nonparametric Levene test for equal variances. *Psicológica*, *31*(2), 401–430. Retrieved [26.5.21] from: https://www.uv.es/psicologica/articulos2.10/11NORDSTOKKE.pdf

OECD (Ed.). (2016). *Equations and inequalities: Making mathematics accessible to all*. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264258495-en.

OECD. (2020). *PISA 2018 Results (Volume II) Where All Students Can Succeed*. Paris: OECD Publishing. https://doi.org/10.1787/b5fd1b8f-en

Official Statistics of Finland. (OSF). 2020. *Providers of education and educational institutions* [e-publication]. ISSN=1799-5825. 2019. Helsinki: Statistics Finland [referred: 15.9.2020]. Retrieved [26.5.2021] from: http://www.stat.fi/til/kjarj/2019/kjarj_2019_2020-02-12_tie_001_en.html

Paquette, D. & Ryan, J. (2001). *Bronfenbrenner's Ecological Systems Theory*. Retrieved [26.5.2021] from: http://dropoutprevention.org/wp-content/uploads/2015/07/paquetteryanwebquest_20091110.pdf

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and and school achievement. In R. J. Riding & S. G. Rayner (Eds.), *International perspectives on individual differences: Vol. 2. Self perception* (pp. 239–265). Ablex Publishing. Retrieved [26.5.2021] from: https://www.uky.edu/~eushe2/Pajares/PajaresSchunk2001.html

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-Group Effects: Instructional, Social, or Institutional? *Sociology of Education*, *67*(1), 27. https://doi.org/10.2307/2112748

Parker, Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*(1), 29–48. https://doi.org/10.1080/01443410.2013.797339

Parker, P., Dicke, T., Guo, J., & Marsh, H. (2019). *A macro context theory of academic self-concept: Ability stratification and the big-fish-little-pond effect* [Preprint]. https://doi.org/10.31234/osf.io/bwy59

Parker, Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology*, *48*(6), 1629–1642. https://doi.org/10.1037/a0029167

Parker, Van Zanden, B., & Parker, R. B. (2018). Girls get smart, boys get smug: Historical changes in gender differences in math, literacy, and academic social comparison and achievement. *Learning and Instruction*, *54*, 125–137. https://doi.org/10.1016/j.learninstruc.2017.09.002

Peetsma, T., van der Veen, I., Koopman, P., & van Schooten, E. (2006). Class composition influences on pupils' cognitive development 1. *School Effectiveness and School Improvement*, *17*(3), 275–302. https://doi.org/10.1080/13803610500480114

Pekrun, R. (2017). Achievement emotions. In Elliot, C. Dweck & D. Yager (Eds.), *Handbook of Competence and Motivation: Theory and Application* (Second edition, pp. 251-271). New York: The Guilford Press. ISBN: 978-1-4625-3603-0

Pintrich, P., & Schunk, D. (1996). The Role of Expectancy and Self-Efficacy Beliefs. In *Motivation in Education: Theory, Research & Applications*. Prentice Hall.

Raigodorosky. (2008). Райгородский, Д.Я. (ed.) (2008). Энциклопедия психодиагностики. Психодиагностика детей, 177-179.

Rawlings, A., Tapola, A., & Niemivirta, M. (2020). Temperamental Sensitivities Differentially Linked with Interest, Strain, and Effort Appraisals. *Frontiers in Psychology*. https://doi.org/10.31234/osf.io/24wc5

Reay, D., & Lucey, H. (2003). The Limits of Choice': Children and Inner City Schooling. *Sociology*, 37(1), 121-142. https://doi.org/10.1177%2F0038038503037001389

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology*, *107*(3), 645–662. https://doi.org/10.1037/edu0000012

Rinne, R., Seppänen, P. & Kosunen, S. (2017). Yläkoulujen valinnat eivät jakaudu tasa-arvoisesti: vastine Timo Jaloselle. *Kasvatus*, 48 (2), 145-147.

Rinne, R. (2000). The Globalisation of Education: Finnish education on the doorstep of the new EU millennium. *Educational Review*, *52*(2), 131–142. https://doi.org/10.1080/713664043

Roque, I., de Lemos, M. S., & Gonçalves, T. (2014). A longitudinal developmental analysis of students' causality beliefs about school performance. *European Journal of Psychology of Education*, *29*(1), 159–173. https://doi.org/10.1007/s10212-013-0192-3

Rosenbaum, P. & Rubin, (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Ross, J. D., & Ross, C. M. (1979). *Ross test of Higher Cognitive Processes.* Academic Therapy Publications.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology*, *16*(2), 105–115. https://doi.org/10.1037/0012-1649.16.2.105

Ryan, A. M. (2000). Peer Groups as a Context for the Socialization of Adolescents' Motivation, Engagement, and Achievement in School. *Educational Psychologist*, *35*(2), 101–111. https://doi.org/10.1207/S15326985EP3502_4

Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the big-fish–little-pond effect across cultures. *Journal of Educational Psychology*, *108*(3), 405–423. https://doi.org/10.1037/edu0000063

Saleh, M., Lazonder, A. W., & De Jong, T. (2005). Effects of within-class ability grouping on social interaction, achievement, and motivation. *Instructional Science*, *33*(2), 105–119. https://doi.org/10.1007/s11251-004-6405-z

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos*, *61*(2), 279–308. https://doi.org/10.1111/j.1467-6435.2008.00402.x

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, *101*(2), 403–419. https://doi.org/10.1037/a0013838

Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., Blanton, H., Buunk, A. P., Gibbons, F. X., Kuyper, H., Suls, J., & Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology*, *47*(1), 73–103. https://doi.org/10.1348/014466607X202309

Seppänen, P. (2003). Patterns of 'public-school markets' in the Finnish comprehensive school from a comparative perspective. *Journal of Education Policy*, *18*(5), 513–531. https://doi.org/10.1080/0268093032000124875

Seppänen, P. (2006). *Kouluvalintapolitiikka perusopetuksessa: Suomalaiskaupunkien koulumarkkinat kansainvälisessä valossa*. Kasvatusalan tutkimuksia 26. Turku: Suomen kasvatustieteellinen seura. ISBN: 952-5402-25-1

Seppänen, P., Carrasco, A., Kalalahti, M., Rinne, R., & Simola, H. (Ed), (2015). *Contrasting dynamics in education politics of extremes: School choice in Chile and Finland*. Comparative and International education: A Diversity of Voices. Vol. 37. Rotterdam: Sense Publishers. ISBN: 978-94-6300-260-8

Seppänen, P., Kalalahti, M., Rinne, R., & Simola, H. (2015). *Lohkoutuva peruskoulu. Perheiden kouluvalinnat, yhteiskuntaluokat ja koulutuspolitiikka. [Segmenting comprehensive school: Parental school choice, social classes and education policies]*. Finnish Educational Research Association. ISBN: 978-952-5401-70-7

Seppänen, P., Rinne, R., & Riipinen, P. (2012). Yläkouluvalinnat, koulujen suosio ja perheiden sosiaalinen asema – Lohkoutuuko perusopetus kaupungeissa? *Kasvatus*, *43*(3), 226–243.

Seppänen, P.& Rinne, R. (2015). Suomalainen yhtenäiskoulu ylikansallisen koulupolitiikan paineissa. In P. Seppänen, M. Kalalahti, R. Rinne & H. Simola (Eds.), *Lohkoutuva peruskoulu. Perheiden kouluvalinnat,yhteiskuntaluokat ja koulutuspolitiikka. [Segmenting comprehensive school: Parental school choice, social classes and education policies]* (pp. 24–58). Finnish Educational Research Association. ISBN: 978-952-5401-70-7

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Silva, B. C., Bosancianu, C. M., & Littvay, L. (2019). *Multilevel structural equation modeling*. Sage. ISBN: 978-1-5443-2305-3

Silvennoinen, H., Kalalahti, M., & Varjo, J. (2015). Why Fix something that is not Broken? The Implementation of School Choice Policy and Parental Attitudes towards Equality and Uniformity of Comprehensive School System in Finland. *Athens Journal of Education*, *2*(1), 37–52. https://doi.org/10.30958/aje.2-1-3

Simola, H., Seppänen, P., Kosunen, S., & Vartiainen, H. (2015). Oppilasvalikoinnin paluu? [Return of pupil selection?]. In P. Seppänen, M. Kalalahti, R. Rinne & H. Simola (Eds.), *Lohkoutuva peruskoulu. Perheiden kouluvalinnat,yhteiskuntaluokat ja koulutuspolitiikka. [Segmenting comprehensive school: Parental school choice, social classes and education policies]* (pp. 87–122). Finnish Educational Research Association. ISBN: 978-952-5401-70-7

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and External Frames of Reference for Academic Self-Concept. *Educational Psychologist*, *37*(4), 233–244. https://doi.org/10.1207/S15326985EP3704_3

Skaalvik, & Skaalvik, E. M. (2004). Gender Differences in Math and Verbal Self-Concept, Performance Expectations, and Motivation. *Sex Roles*, *50*(3/4), 241–252. https://doi.org/10.1023/B:SERS.0000015555.40976.e6

Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology*, *71*(3), 549–570. https://doi.org/10.1037/0022-3514.71.3.549

Skinner, E. A., Chapman, M., & Baltes, P. B. (1988). Control, Means-Ends, and Agency Beliefs: A New Conceptualization and Its Measurement During Childhood. *Journal of Personality and Social Psychology*, *54*(1), 117–133. https://doi/10.1037/0022-3514.54.1.117

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What It Takes to Do Well in School and Whether I've Got It: A Process Model of Perceived Control and Children's Engagement and Achievement in School. *Journal of Educational Psychology*, *82*(1), 22–32. https://doi/10.1037/0022-0663.82.1.22

Skinner, E. A., Zimmer-Gembeck, M. J., Connell, J. P., Eccles, J. S., & Wellborn, J. G. (1998). *Individual differences and the development of perceived control.* Monographs of the society for research in child development. Vol.63, No. 2/3. Wiley. https://doi.org/10.2307/1166220

SPSS Inc. (2005). *Linear Mixed-Effects Modeling in SPSS: An Introduction to the MIXED Procedure* [Technical report]. Retrieved [26.5.2021] from: https://csass.ucsc.edu/self-study%20courses/SPSSMixed.pdf

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses. *Review of Educational Research*, *86*(4), 849–899. https://doi.org/10.3102%2F0034654316675417

Stetsenko, A., Little, T. D., Gordeeva, T., Grasshof, M., & Oettingen, G. (2000). Gender Effects in Children's Beliefs about School Performance: A Cross-Cultural Study. *Child Development*, *71*(2), 517–527. https://doi.org/10.1111/1467-8624.00161

Stipek, D., & MacIver, D. (1989). Developmental Change in Children's Assessment of Intellectual Competence. *Child Development*, *60*(3), 521–538. https://doi.org/10.2307/1130719

Stipek, & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of Educational Psychology*, *83*(3), 361–371. https://doi.org/10.1037/0022-0663.83.3.361

Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: Consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, *26*(1), 75–101. https://doi.org/10.1080/09243453.2013.871302

TENK. (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland. Guidelines of the Finnish Advisory Board on Research Integrity 2012*. Finnish Advisory Board on Research Integrity TENK. Retrieved [26.5.2021] from: https://www.tenk.fi/en/tenk-guidelines

Thijs, J., Verkuyten, M., & Helmond, P. (2010). A Further Examination of the Big-Fish–Little-Pond Effect: Perceived Position in Class, Class Size, and Gender Comparisons. *Sociology of Education*, *83*(4), 333–345. https://doi.org/10.1177/0038040710383521

Tieso, C. L. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, *26*(1), 29–36. https://doi.org/10.1080/02783190309554236

Trautwein, U., Köller, O., Lüdtke, O., & Baumert, J. (2005). Student tracking and the powerful effects of opt-in courses on self-concept: Reflected-glory effects do exist after all. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *New frontiers for self research* (pp. 307-327). Greenwich, Ct.: Information Age Publishing.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, *101*(4), 853–866. https://doi.org/10.1037/a0016306

Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology*, *44*(5), 331–349. https://doi.org/10.1016/j.jsp.2006.04.003

Usher, E. L., & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research*, *78*(4), 751–796. https://doi.org/10.3102/0034654308321456

Vainikainen. (2014). *Finnish primary school pupils' performance in learning to learn assessments: A longitudinal perspective on educational equity*. Research Report 360. Helsinki: University of Helsinki. ISBN 978-951-51-0185-3

Vainikainen, M.-P., Wüstenberg, S., Kupiainen, S., Hotulainen, R., & Hautamäki, J. (2015). Development of learning to learn skills in primary school. *International Journal of Lifelong Education*, *34*(4), 376–392. https://doi.org/10.1080/02601370.2015.1060025

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, *5*(2), 134–150. https://doi.org/10.1016/j.edurev.2010.02.001

Van Houtte, M. (2016). Lower-track students' sense of academic futility: Selection or effect? *Journal of Sociology*, *52*(4), 874–889. https://doi.org/10.1177/1440783315600802

Van Houtte, M., & Stevens, P. A. J. (2015). Tracking and sense of futility: The impact of between-school tracking versus within-school tracking in secondary education in Flanders (Belgium). *British Educational Research Journal*, *41*(5), 782–800. https://doi.org/10.1002/berj.3172

Varjo, J., & Kalalahti, M. (2011). Koulumarkkinoiden institutionaalisen tilan rakentuminen. *Yhdyskuntasuunnittelu*, *49*(4). Retrieved [26.5.2021] from: http://www.yss.fi/varjo_kalalahti.pdf

Varjo, J., & Kalalahti, M. (2015). The conceivable benefits of being comprehensive – Finnish Local Education Authorities on recognising and controlling the social costs of school choice. *European Educational Research Journal*, *14*(3–4), 312–330. https://doi.org/10.1177/1474904115589036

Varjo, J., & Kalalahti, M. (2019). The art of governing local education markets – municipalities and school choice in Finland. *Education Inquiry*, *10*(2), 151–165. https://doi.org/10.1080/20004508.2018.1514907

Varjo, J., Kalalahti, M., & Silvennoinen, H. (2014). Families, School Choice, and Democratic Iterations on the Right to Education and Freedom of Education in Finnish Municipalities. *Journal of School Choice*, *8*(1), 20–48. https://doi.org/10.1080/15582159.2014.875408

Vettenranta, J., Välijärvi, J., Ahonen, A., Hautamäki, J., Hiltunen, J., Leino, K., Lähteinen, S., Nissinen, K., Puhakka, E., Rautopuro, J., & Vainikainen, M.-P. (2016). *PISA 15 Ensituloksia: Huipulla pudotuksesta huolimatta [PISA 15 first reports: Excellent despite the drop]*. Opetus- ja kulttuuriministeriön julkaisuja 2016:41. Helsinki: Opetus ja kulttuuriministeriö.

Vincent, C. (2001). Social class and parental agency. *Journal of Education Policy*, *16*(4), 347–364. https://doi.org/10.1080/0268093011-54344

Vincent, C., & Ball, S. J. (2007). `Making Up' the Middle-Class Child: Families, Activities and Class Dispositions. *Sociology*, *41*(6), 1061–1077. https://doi.org/10.1177/0038038507082315

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Vygotsky, L. S. (1978): Mind in Society. The Development of Higher Psychological Processes. Edited by M. Cole, V. John-Steiner, S. Scribner & E. Souberman. Cambridge, Massachusetts, Harvard University Press. ISBN: 978-0-6745-7629-2

Wallenius, T. (2020). Schools, Performance and Publicity : Contrasting the policy on publicising school performance indicators in Finland with the other Nordic countries. *Helsinki Studies in Education* 92. Helsinki: Unigrafia. ISBN 978-951-51-6451-3

Walls, T. A., & Little, T. D. (2005). Relations Among Personal Agency, Motivation, and School Adjustment in Early Adolescence. *Journal of Educational Psychology*, *97*(1), 23–31. https://doi.org/10.1037/0022-0663.97.1.23

Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118356258

Wechsler, D. (1981). *WAIS-R: Manual: Wechsler Adult Intelligence Scale – revised.* Harcourt Brace Jovanovich for Psychological Corp.

Weiner, B. (1985). An Attributional Theory of Achievement Motivation and Emotion. *Psychological Review, 92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Wentzel, K. R. (2017). Peer Relationhips, Motivation and Academic Performance at School. In A. Elliot, C. Dweck & D. Yager (Eds.), *Handbook of competence and motivation: Theory and application* (Second edition, pp. 586–603). New York: The Guilford Press. ISBN: 978-1-4625-3603-0

Wentzel, K. R., & Caldwell, K. (1997). Friendships, Peer Acceptance, and Group Membership: Relations to Academic Achievement in Middle School. *Child Development*, *68*(6), 1198. https://doi.org/10.2307/1132301

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*(5), 297–332. https://doi.org/10.1037/h0040934

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. *Child Development Perspectives*, *4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Willms, J. D. (1985). The Balance Thesis: Contextual effects of ability on pupils' O-grade examination results. *Oxford Review of Education*, *11*(1), 33–41. https://doi.org/10.1080/0305498850110103

Wilson, A. E., & Ross, M. (2000). The frequency of temporal-self and social comparisons in people's personal appraisals. *Journal of Personality and Social Psychology*, *78*(5), 928–942. https://doi.org/10.1037/0022-3514.78.5.928

Wilson, Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, *5*(3), 249–255. https://doi.org/10.1111/j.2044-835X.1987.tb01060.x

Wößmann, L. (2009). *International evidence on school tracking: A review*. 9. Retrieved [26.5.2021] from: https://www.ifo.de/DocDL/dicereport109-rr1.pdf

Wouters, S., Colpin, H., Van Damme, J., De Laet, S., & Verschueren, K. (2013). Early adolescents' academic self-concept formation: Do classmates or friends matter most? *Learning and Individual Differences*, *27*, 193–200. https://doi.org/10.1016/j.lindif.2013.09.002

Wright, D. B., London, K., & Field, A. P. (2011). Using Bootstrap Estimation and the Plug-in Principle for Clinical Psychology Data. *Journal of Experimental Psychopathology*, *2*(2), 252–270. https://doi.org/10.5127/jep.013611

Yang Hansen, K., Gustafsson, J., & Rosen, M. (2014). School performance differences and policy variations in Finland, Norway and Sweden. In T. Hallvard (Eds.), *Northern Lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries* (pp. 25–48). Nordic Council of Ministers. Retrieved [26.5.2021] from: https://www.uv.uio.no/ils/forskning/prosjekter/timss/publikasjoner/northernlightstimss-pirls.pdf

Ylönen, A. (2009). *Specialisation within the Finnish comprehensive school system: Reasons and outcomes for equity and equality of opportunity*. VDM Verlag Dr. Müller. ISBN: 978-3-639-17885-2

Zell, E., & Alicke, M. D. (2010). The Local Dominance Effect in Self-Evaluation: Evidence and Explanations. *Personality and Social Psychology Review*, *14*(4), 368–384. https://doi.org/10.1177/1088868310366144

Zimmer, R. (2003). A new twist in the educational tracking debate. *Economics of Education Review*, *22*(3), 307–315. https://doi.org/10.1016/S0272-7757(02)00055-9

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology*, 81(3), 329. https://psycnet.apa.org/doi/10.1037/0022-0663.81.3.329

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, *57*(1), 173–181. https://doi.org/10.1348/000711004849222

Appendices

Appendix A. Number of same respondents by scales in each year

Number of participants in different measurement points (Years)							
Scale	Only Year 4	Only Year 6	Both years 4 and 6	Only year 7	Only year 9	Both years 7 and 9	All four years
Agency: Ability	953	969	844	2513	3518	1965	633
Agency: Effort	954	960	838	2395	3099	1687	557
CE: Success	964	970	850	2519	3527	1974	634
CE: Learning	-	-	-	2398	3113	1691	-
ME: Luck	-	-	-	2394	3058	1663	-
ME: Ability	-	-	-	2392	3110	1691	-
ASC: Math	963	970	850	2515	3525	1970	633

Appendix B. Correlations among variables in sub studies

Correlations between variables in each sub study are presented in tables below. Statistically significant correlations ($p < .05$) are presented in bold and non-significant correlations are in gray.

Correlations among variables, study I

Variables	*N*	1	2	3	4	5	6	7	8	9	10
1. Agency: Ability Year 4	947	**1.00**									
2. Agency: Ability Year 6	960	**.28**	**1.00**								
3. Agency: Effort Year 4	938	-	-	**1.00**							
4. Agency: Effort Year 6	944	-	-	-	**1.00**						
5. Control expectancy: Success Year 4	957	-	-	-	-	**1.00**					
6. Control expectancy: Success Year 6	965	-	-	-	-	-	**1.00**				
7. Mother's education	781	**.10**	**.21**	.07	**.10**	**.16**	**.19**	**1.00**			
8. Class type	1025	.00	**.10**	.02	.04	.04	**.08**	**.27**	**1.00**		
9. Prior achievement	961	**.28**	**.36**	**.17**	**.15**	**.27**	**.28**	**.32**	**.10**	**1.00**	
10. Gender	965	.02	**.10**	**.14**	**.10**	.04	.04	.00	**-.10**	**-.07**	**1.00**

NOTE: Correlations between different dimensions of action-control beliefs were not included as analyses were run separately for each dimension.

Correlations among variables, study II

Variables	*N*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Agency: Ability Year 7	1029	**1.00**															
2. Agency: Ability Year 9	1029	**.59**	**1.00**														
3. Agency: Effort Year 7	1008	**.57**	**.40**	**1.00**													
4. Agency: Effort Year 9	1008	**.37**	**.49**	**.58**	**1.00**												
5. CE: Success Year 7	1029	**.70**	**.49**	**.44**	**.29**	**1.00**											
6. CE: Success Year 9	1029	**.49**	**.77**	**.35**	**.40**	**.55**	**1.00**										
7. CE: Learning Year 7	1009	**.67**	**.44**	**.68**	**.40**	**.66**	**.45**	**1.00**									
8. CE: Learning Year 9	1009	**.47**	**.63**	**.44**	**.60**	**.40**	**.63**	**.54**	**1.00**								
9. MEB: Luck Year 7	1008	**-.28**	**-.25**	**.13**	**-.11**	**-.17**	**-.18**	**-.14**	**-.16**	**1.00**							
10. MEB: Luck Year 9	1008	**-.19**	**-.17**	**.08**	-.01	**-.11**	**-.14**	**-.07**	**-.11**	**.48**	**1.00**						
11. MEB: Ability Year 7	1008	**-.17**	**-.19**	**.50**	-.04	**.12**	**-.13**	-.03	**-.10**	**.54**	**.31**	**1.00**					
12. MEB: Ability Year 9	1008	**-.12**	**-.15**	**.27**	**.08**	**-.08**	**-.08**	-.03	.02	**.28**	**.52**	**.53**	**1.00**				
13. Mother's education	975	**.18**	**.21**	**.12**	**.11**	**.18**	**.18**	**.14**	**.17**	**-.12**	**-.08**	**-.13**	-.02	**1.00**			
14. Class type	1029	**.12**	**.10**	**.09**	**.10**	**.17**	**.10**	**.13**	**.13**	**-.10**	.01	-.05	.03	**.15**	**1.00**		
15. Prior achievement	999	**.30**	**.32**	**.17**	**.12**	**.23**	**.25**	**.19**	**.25**	**-.36**	**-.21**	**-.31**	**-.15**	**.31**	**.16**	**1.00**	
16. Gender	1007	**.07**	**.04**	**.10**	-.04	**.08**	.06	**.07**	.03	**.12**	**.18**	**.22**	**.224**	.00	**.10**	-.06	**1.00**

NOTE: CE refers to Control expectancy and MEB to Means-end-beliefs. Matrix is based on the randomized sample, where group sizes were equalized before the analyses, therefore the number of respondents is smaller than in the original sample (see more study II)

Correlations among variables study III											
Variables	*N*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. Mathematical Self-concept Year 4	939	**1.00**									
2. Mathematical Self-concept Year 6	882	**.48**	**1.00**								
3. Mathematical thinking skills Year 4	998	**.34**	**.35**	**1.00**							
4. Mathematical thinking skills Year 6	892	**.13**	**.14**	**.41**	**1.00**						
5. Mother's education	781	**.10**	**.14**	**.26**	**.30**	**1.00**					
6. Class type	939	.03	.06	**.13**	**.15**	**.27**	**1.00**				
7. Gender	1039	**.21**	**.20**	**.15**	**.07**	.02	**-.07**	**1.00**			
8. Class-average math skills Year 4	1028	**.11**	**.12**	**.36**	**.28**	**.33**	.02	**.28**	**1.00**		
9. Mother's education ratio class level	1022	-.02	.04	**.28**	**.30**	**.51**	**.50**	.00	**.66**	**1.00**	
10. Gender ratio class level	1028	**.10**	-.02	.03	-.01	-.02	**-.33**	**.25**	.06	-.02	**1.00**

Correlations among variables, study IV

Variables	*N*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. Mathematical Self-concept Year 4	945	**1.00**								
2. Mathematical Self-concept Year 6	881	**.36**	**1.00**							
3. Mathematical Test Year 4	943	**.30**	**.32**	**1.00**						
4. Teacher Rating of Math Achievement Year4	945	**.32**	**.30**	**.42**	**1.00**					
5. Gender (1 = girl, 2 = boy)	1004	**.18**	**.18**	**.07**	**.09**	**1.00**				
6. Highest parental educational level	824	-.02	**.09**	**.14**	**.11**	.02	**1.00**			
7. School Readiness Test Year 1	602	.07	**.09**	**.15**	**.12**	.02	-.03	**1.00**		
8. Class-average mathematical test Year 4	1017	-.02	.05	**.43**	**.14**	.01	**.23**	.02	**1.00**	
9. Peer-group-average mathematical test Year 4	1017	.05	**.11**	**.53**	**.21**	**.13**	**.23**	.03	**.80**	**1.00**

Appendix C. Descriptive information of items

Item	Year	min	max	M	SD	skew./std.error	kurt./sted.error	N
Agency: ability								
I am clever enough to do well at school	4	1	7	5.81	1.33	-1.324/.079	1.910/.159	947
	6	1	7	5.77	1.21	-1.024/.079	.892/.158	956
	7	1	7	5.75	1.21	-1.093/-058	1.313/.115	1807
	9	1	7	5.63	1.38	-1.132/.057	1.134/.115	1818
I am a clever and able student	4	1	7	5.68	1.35	-1.245/.080	1.772/.159	944
	6	1	7	5.11	1.38	-.711/.079	.298/.158	959
	7	1	7	5.22	1.37	-.778/.058	.475/.115	1795
	9	1		5.17	1.48	-.831/.057	.367/.115	1814
I have the abilities necessary for success at school	4	1	7	5.95	1.32	-1.605/.080	2.832/.159	942
	6	1	7	5.87	1.16	-1.131/.079	1.331/.158	960
	7	1	7	5.85	1.17	-1.137/.058	1.388/.115	1800
	9	1	7	5.72	1.34	-1.151/.057	1.104/.115	1813
Agency: effort								
I try enough at school	4	1	7	5.89	1.30	-1.509/.080	2.523/.160	938
	6	1	7	5.61	1.32	-1.041/.079	.914/.159	950
	7	1	7	5.52	1.28	-.803/.059	.547/.119	1704
	9	1	7	4.99	1.66	-.648/.062	-.190/.124	1567
I work hard to do well at school	4	1	7	5.54	1.46	-1.140/.080	1.088/.159	944
	6	1	7	4.97	1.37	-.450/.079	-.056/.158	951
	7	1	7	5.09	1.33	-.493/.059	.043/.119	1704
	9	1	7	4.66	1.54	-.387/.062	-.262/.124	1567
I concentrate well enough in class	4	1	7	5.88	1.24	-1.482/.080	2.587/.159	941
	6	1	7	5.27	1.34	-.661/.080	.086/.159	943
	7	1	7	5.32	1.29	-.627/.059	.144/.118	1718
	9	1	7	4.91	1.54	-.576/.061	-.105/.123	1594
CE: success								
I can get good marks at school if I want to	4	1	7	4.69	1.81	-.636/.079	-.484/.158	957
	6	1	7	4.94	1.52	-.653/.079	-.084/.157	965
	7	1	7	5.14	1.45	-.713/.058	.148/.115	1811
	9	1	7	5.20	1.50	-.737/.057	.020/.115	1822
I can learn the things required to if I decide to	4	1	7	5.36	1.59	-.950/.080	.350/.159	940
	6	1	7	5.54	1.31	-.906/.079	.532/.158	958
	7	1	7	5.53	1.311	-1.005/.058	1.055/.115	1804
	9	1	7	5.41	1.40	-.945/.057	.653/.115	1816
I can do well at school if I decide to	4	1	7	5.17	1.61	-.824/.080	.118/.160	935
	6	1	7	5.43	1.34	-.946/.079	.816/.158	957
	7	1	7	5.52	1.311	-.888/.058	.574/.115	1801
	9	1	7	5.31	1.49	-.864/.057	.305/.115	1817
CE: learning								
I can learn things taught at school, if I want	7	1	7	5.58	1.22	-.879/.059	1.003/.118	1729
	9	1	7	5.23	1.47	-.816/.061	.377/.122	1611
If I don't yet master something I need to, I'll learn it easily enough.	7	1	7	4.97	1.36	-.415/.059	-.170/.118	1716
	9	1	7	4.71	1.530	-.457/.061	-.185/.123	1589
If learning something is important to me, I know I'll learn it.	7	1	7	5.46	1.30	-.729/.059	.374/.118	1706
	9	1	7	5.27	1.49	-.760/.062	.259/.124	1568

ME: luck								
Success at school is a matter of luck.	7	1	7	2.45	1.70	1.061/.059	.160/.118	1713
	9	1	7	2.52	1.7	.996/.061	.109/.123	1585
One cannot really influence one's success at school.	7	1	7	2.34	1.86	1.269/.059	.397/.118	1716
	9	1	7	2.43	1.78	1.113/.061	.171/.123	1589
Failure at school is mainly due to bad luck.	7	1	7	3.01	1.80	.584/.059	-.668/.118	1718
	9	1	7	2.90	1.70	.716/.061	-.218/.123	1591
ME: ability								
If one fails at school, it just shows that one is not smart enough.	7	1	7	3.52	1.74	.222/.059	-.805/.118	1706
	9	1	7	3.40	1.71	.388/.062	-.571/.124	1562
Poor marks are due to lack of ability.	7	1	7	3.78	1.72	.078/.059	-.816/.118	1725
	9	1	7	3.52	1.71	.310/.061	-.646/.122	1615
Abilities are necessary for success at school.	7	1	7	4.32	1.54	-.221/.059	-.347/.118	1708
	9	1	7	4.11	1.62	-.124/.061	-.469/.123	1586
AS: mathematics								
Mathematics is very easy for me	4	1	7	5.11	1.60	-.673/.080	-.167/.159	939
	6	1	7	4.96	1.54	-.657/.079	-.066/.158	961
I usually handle even the more difficult math problems well	4	1	7	5.10	1.63	-.774/.080	-.027/.160	936
	6	1	7	4.62	1.62	-.502/.079	-.440/.158	958
I am good in mathematics	4	1	7	5.36	1.66	-.1008/.080	.294/.160	932
	6	1	7	4.70	1.75	-.599/.079	-.526/.158	945

Appendix D. Intraclass correlations of the sum scores

Intraclass correlations (ICC)				
Scale	Year 4	Year 6	Year 7	Year 9
Agency: Ability	.030	.031	.040	.032
Agency: Effort	.012	.049	.030	.048
CE: Success	.034	.016	.023	.029
CE: Learning	-	-	.034	.043
ME: Luck	-	-	.085	.006
ME: Ability	-	-	.047	.010
ASC: Mathematics	.018	.012	-	-
Mathematical thinking skills	.162	.157	-	-

NOTE: Intraclass correlations are calculated for the sum scores with Mplus analysis "twolevel basic"-option. At the primary school data they are based on year 4 class code as a cluster and at the lower secondary school on year 7 class code as a cluster.